

# A STATISTICAL EQUILIBRIUM PERSPECTIVE ON CORPORATE PROFITABILITY

PHILIPP MUNDT

Doctoral thesis

Submitted to the Faculty of Social Sciences, Economics, and Business  
Administration  
University of Bamberg



and to the Faculty of Law and Economics  
University of Castellon



February 2017

Dissertation committee:  
Prof. Mishael Milaković, PhD  
Prof. Dr. Frank Westerhoff  
Prof. Dr. Simone Alfarano



To my parents, Heidemarie and Herbert.





## PREFACE

---

My doctoral thesis consists of four single papers (chapters) which are intended for publication in academic journals. Although all these papers are related and cover different aspects of a common theme, each of them can be read on its own and does not require any prior knowledge from the other chapters. It naturally follows that this style of presentation results in some overlap between the chapters, e.g. with respect to the core model, data description, and related literature. The readers of my entire dissertation will hopefully excuse the presence of potential redundancies that cannot be avoided at this point.

Up to the date of submission of this dissertation in February 2017, two out of four papers have been published in international peer-reviewed journals. The article "Gibrat's law redux: think profitability instead of growth", co-authored with Mishael Milaković and Simone Alfarano, was published in *Industrial and Corporate Change*, 2016, Vol. 25, No. 4, 549-571. The second article "The real versus the financial economy: a global tale of stability versus volatility", co-authored with Niels Förster, Simone Alfarano, and Mishael Milaković, was published in *Economics: The Open-Access, Open-Assessment E-Journal*, 2014, Vol. 8, 2014-17. The two remaining papers "The ultimate corporate objective is survival" and "A statistical equilibrium approach to forecasting corporate profitability" are unpublished to date. In one way or another, all four essays revolve around statistical regularities in the dynamics of business firms.

My interest in the behavior of large, publicly traded corporations rests on the granular view of aggregate fluctuations that has received growing attention in recent years (e.g. Acemoglu et al., 2012; Carvalho and Gabaix, 2013; Gabaix, 2011). The fundamental idea of this literature is that the dynamics of large corporates may be of huge quantitative importance for macroeconomic quantities, e.g. GDP volatility, implying that a better understanding of these dynamics may contribute to a kind of new "micro-foundation" of macroeconomics that is empirically sound and well-founded. Although macroeconomic insights are beyond the scope of this dissertation, I would hope that the identification of empirical regularities in firm dynamics is a first step towards such a theory.

To explain how these four papers are related and why they hopefully contribute to a larger whole, it is instructive to add some remarks about the underlying vision and research methodology. At the most fundamental level, the common theme underlying all papers is the problem of aggregation that economic theory often circumvents with the representative agent paradigm. If agents were homogeneous, it would be sufficient to consider the behavior of a single agent and the aggregate properties of the system could be derived from the behavior of this individual. However, research originating in the field of statistical physics suggests that this reasoning is misleading in complex systems that are characterized by a large number of heterogeneous

interacting agents, or as Aoki and Yoshikawa (2007) put it (p. 26), “Micro behaviors of the representative agent do not mimic the behavior of the macroeconomy. Macroeconomic phenomena are the outcomes of interactions of a large number of economic agents such as households and firms.” Statistical physicists developed an alternative methodology to describe these systems. Their approach builds on the perception that complex dynamics and interactions among the system’s constituent units may produce robust statistical regularities at a higher level of aggregation that do not directly reflect the behavior of the subsystems. Following this idea, we need to shift the focus from a deterministic to a probabilistic modeling approach that takes into account the presence of fluctuations (even in equilibrium), or to quote Aoki and Yoshikawa (2007) again (p. 26), “Equilibrium in the macroeconomy is better described by a probability distribution than by a ‘point’ in some space or set.” Thus, the present collection of articles approaches the dynamics of business firms from a probabilistic perspective.

To detect these macro-regularities, one needs to select suitable quantities to characterize the states the system may take at every point in time. Browsing the literature that has been published in the research field of industrial dynamics so far (part of which is quoted in the first chapter), one is left with the impression that most papers in this vein focus on the growth rate of firm size. To this end, the first chapter of this dissertation explores the cross-sectional and time series properties of profit and growth rates in closer detail to show that profit rates are economically more fundamental and statistically more convenient than growth rates, and that Robert Gibrat’s seminal idea of a common law governing the dynamics of all firms applies to profitability but not firm growth.

The paper presented in the second chapter replicates some of the results pertaining to the statistical properties of profit and growth rates on a much broader basis considering data of about 30,000 firms from more than 40 countries. It shows that the stability of the average profit rate and its volatility is not a peculiarity of the US data but a rather universal feature that can be observed across countries. Defining firm size in terms of market value, this paper also contributes to the literature stressing “excess volatility” in financial returns.

The third chapter investigates the impact of firm idiosyncrasies on profitability. Given that profit rates are in statistical equilibrium, one should expect that profit rates are governed by a common probabilistic law of motion for all firms. Such a view is diametrically opposed to previous findings in the industrial organization, strategic management, and accounting and finance literature which stress the importance of sectoral or firm-level effects on performance. It turns out that survival time may be a potential explanation for the emergence of such divergent views because a common law for profitability only prevails conditional on survival. This leads to the fundamental question of what the ultimate business objective should be.

Finally, the fourth chapter presents an application of the statistical equilibrium methodology to forecasting profitability.

Overall, I think that the results presented in this dissertation suggest that statistical equilibrium is a very reasonable first order approximation to the profitability of surviving corporations, and that methods for the modeling of complex systems do have

explanatory power for problems in economics. One of the major questions for future research should be if these insights on the dynamics of large, long-lived corporations can be exploited to draw inferences on the macroeconomy.



## PREFACIO

---

Mi tesis doctoral consta de cuatro artículos (capítulos) que están destinados a ser publicados en revistas académicas. Aunque todos estos artículos están relacionados, en el sentido de que abarcan diferentes aspectos de un tema común, cada uno de ellos puede leerse por sí solo y no requiere ningún conocimiento previo de los otros capítulos. Debería deducirse con bastante naturalidad que este estilo de presentación da lugar a una cierta superposición entre los capítulos, por ejemplo con respecto al modelo básico, la descripción de los datos y la literatura relacionada. Dicho esto, ha de entenderse y esperarse que los lectores de mi tesis disculpen la presencia de redundancias potenciales inevitables en este momento.

Hasta la fecha de presentación de esta tesis en febrero de 2017, dos de los cuatro artículos han sido publicados en revistas indexadas internacionales. El artículo “Gibrat’s law redux: think profitability instead of growth”, co-escrito con Mishael Milaković y Simone Alfarano, fue publicado en *Industrial and Corporate Change*, 2016, Vol. 25, No. 4, 549-571. El segundo artículo “The real versus the financial economy: a global tale of stability versus volatility”, en coautoría con Niels Förster, Simone Alfarano y Mishael Milaković, fue publicado en *Economics: The Open-Access, Open-Assessment E-Journal*, 2014, Vol. 8, 2014-17. Los dos artículos restantes “The ultimate corporate objective is survival” y “A statistical equilibrium approach to forecasting corporate profitability” no se han publicado hasta la fecha. De una manera u otra, los cuatro ensayos giran en torno a regularidades estadísticas en la dinámica de las empresas. Complementan y añaden más apoyo empírico al trabajo de Alfarano y Milaković (2008) y Alfarano y col. (2012) que ya sugiere que la distribución de las tasas de beneficio en firme pueden caracterizarse como un resultado estadístico de equilibrio.

Mi interés en el comportamiento de las grandes corporaciones que cotizan en bolsa se funda profundamente en la visión granular de las fluctuaciones agregadas que ha recibido creciente atención en los últimos años (por ejemplo Acemoglu y col., 2012; Carvalho y Gabaix, 2013; Gabaix, 2011). La idea fundamental de esta literatura es que la dinámica de las grandes corporaciones puede ser de gran importancia cuantitativa para cantidades macroeconómicas, por ejemplo la volatilidad del PIB, lo que implica que una mejor comprensión de estas dinámicas puede contribuir a una nueva “micro-fundación” de la macroeconomía empíricamente sólida y bien fundamentada. Aunque las ideas macroeconómicas están más allá del alcance de esta tesis, pretendo que la identificación de las regularidades empíricas en la dinámica de las empresas sea un primer paso hacia una teoría de este tipo.

Para explicar cómo estos cuatro artículos están relacionados y por qué esperamos contribuir a un todo más grande, creo que es instructivo dar un paso atrás y añadir algunas observaciones sobre la visión subyacente y la metodología de investigación. En el nivel más fundamental, el tema común subyacente a todos los artículos es el problema de la agregación en la economía que la teoría económica suele tratar de elu-

dir con el paradigma del agente representativo: si los agentes son homogéneos, basta con considerar el comportamiento de un solo agente y las propiedades agregadas del sistema pueden derivarse del comportamiento de este individuo. Sin embargo, la investigación que se origina en el campo de la física estadística sugiere que tal razonamiento ya no se sostiene en los sistemas complejos que se caracterizan por un gran número de agentes interactivos heterogéneos, o como ya expresó Aoki y Yoshikawa (2007)(p.26): “Los micro comportamientos del agente representativo no imitan el comportamiento de la macroeconomía. Los fenómenos macroeconómicos son el resultado de las interacciones de un gran número de agentes económicos, como los hogares y las empresas.” Afortunadamente, los físicos estadísticos no sólo se dieron cuenta de que las interacciones hacen inviable el análisis de los destinos individuales, sino que también han desarrollado una metodología alternativa para describir estos sistemas. Su enfoque se basa en la percepción de que las dinámicas complejas y las interacciones entre las unidades constituyentes del sistema pueden producir regularidades estadísticas robustas a un nivel más alto de agregación que no reflejan directamente el comportamiento de los subsistemas. Siguiendo esta idea, necesitamos cambiar el foco desde un enfoque determinista a un modelado probabilístico que tenga en cuenta la presencia de fluctuaciones (incluso en equilibrio), o citar Aoki y Yoshikawa (2007) nuevamente (p. 26): “El equilibrio en la macroeconomía está mejor descrito por una distribución de probabilidad que por un ‘punto’ en algún espacio o conjunto.” Por lo tanto, abordaré la dinámica de las empresas desde una perspectiva probabilística.

Para detectar tales macro-regularidades es necesario seleccionar una cantidad adecuada para caracterizar los estados que el sistema puede tomar en cada momento. Examinando la literatura que se ha publicado hasta ahora en el campo de la “dinámica industrial” (parte de la cual se cita en el primer capítulo), da la impresión de que la mayoría de los artículos se centran en la tasa de crecimiento del tamaño de la empresa. Para ello, el primer capítulo de esta tesis explora las propiedades transversales y de series temporales de las tasas de beneficio y crecimiento de la manera más detallada para mostrar que las tasas de beneficios son económicamente más fundamentales y estadísticamente más convenientes que las tasas de crecimiento, y que la idea seminal de Robert Gibrat de una ley común que gobierna la dinámica de todas las empresas se aplica a la rentabilidad, pero no al crecimiento de la empresa.

El artículo presentado en el segundo capítulo reproduce algunos de los resultados relativos a las propiedades estadísticas de las tasas de beneficio y crecimiento sobre una base mucho más amplia considerando datos de unas 30.000 empresas de más de 40 países. Muestra que la estabilidad de la tasa de ganancia promedio y su volatilidad no son peculiaridades de los datos de los Estados Unidos, sino más bien características universales que pueden observarse entre países. Definiendo el tamaño de la empresa en términos de valor de mercado, este artículo también contribuye a la literatura enfatizando la “volatilidad excesiva” en los mercados financieros.

El tercer capítulo investiga el impacto de las idiosincrasias de las empresas en la rentabilidad. Dado que las tasas de beneficio están en equilibrio estadístico, uno debería esperar que las tasas de beneficio se rigen por una ley probabilística de movimiento común para todas las empresas. Este punto de vista es diametralmente opuesto a los

hallazgos previos en organización industrial, gestión estratégica y literatura de contabilidad y finanzas que subrayan la importancia de los efectos sectoriales o a nivel de empresa sobre el rendimiento. Resulta que el tiempo de supervivencia puede ser una posible explicación para la aparición de opiniones tan divergentes, ya que una ley común para la rentabilidad sólo prevalece condicionada a la supervivencia, lo que lleva a la pregunta fundamental de cuál debería ser el objetivo final del negocio.

Finalmente, el cuarto capítulo presenta una aplicación de la metodología de equilibrio estadístico para predecir la rentabilidad.

En general, creo que los resultados presentados en esta tesis sugieren que el equilibrio estadístico es una aproximación de primer orden muy razonable a la rentabilidad de las corporaciones supervivientes y que los métodos para el modelado de sistemas complejos pueden tener un poder explicativo para los problemas económicos. Una de las principales preguntas para la investigación futura debería ser si estos conocimientos sobre la dinámica de las grandes empresas de larga vida pueden ser explotados para extraer inferencias sobre la macroeconomía.





## ACKNOWLEDGEMENTS

---

First and foremost, I would like to thank my supervisors, Mishael Milaković and Simone Alfarano, for their valuable guidance and continuing support in many ways during the course of my studies. Their education shaped my way of thinking about economics, and without them this work could not have been undertaken.

I am also grateful to Frank Westerhoff for his advice as well as for serving on my dissertation committee, Thomas Lux for his ongoing support and help, and to Eva Camacho Cuenca, Miguel Ginés Vilar, and their colleagues from Universitat Jaume I for their hospitality during several research visits in Castellón de la Plana, Spain, where I made substantial progress in my studies.

This work greatly benefited from fruitful discussions with Christian Babirat, Niels Förster, and Ilfan Oh. Ilfan's company during these years has been a privilege, and I will always remain thankful for his friendship and the intellectual stimuli provided by him. Moreover, I am indebted to Laura Linhardt, Laura Louzil, Christian Menden, Lisa Planer-Friedrich, Jan Schulz, and Anja Volk for their able research assistance, and to Roswitha Albert for her help in many ways.

My special thanks go to Derya for all the love, affection, and encouragement, and to my parents for supporting me from Day One. My gratitude for their advice, patience, and unconditional love is beyond expression. I would like to dedicate this dissertation to them.



# CONTENTS

---

1		
	GIBRAT'S LAW REDUX: THINK PROFITABILITY INSTEAD OF GROWTH	1
1.1	Introduction	2
1.2	Data	5
1.2.1	Empirical densities	7
1.2.2	Autocorrelations	9
1.2.3	Size (in)dependence	11
1.3	Model	14
1.3.1	Diffusion	14
1.3.2	Transient density	15
1.3.3	Autocorrelation function	16
1.4	Results	17
1.4.1	Estimation of the diffusion coefficient	17
1.4.2	What determines the diffusion coefficient?	19
1.4.3	Firm size	19
1.4.4	Diversification	20
1.4.5	Intensity of capital	22
1.5	Discussion	22
A	APPENDIX	25
A.1	Sample characteristics	25
A.2	Estimates of the Subbotin parameters	28
A.3	Disaggregation	28
A.4	Simulations	31
A.5	Regression results	33
2		
	THE REAL VERSUS THE FINANCIAL ECONOMY: A GLOBAL TALE OF STABILITY VERSUS VOLATILITY	35
2.1	Introduction	36
2.2	Data description and sample selection	38
2.3	Empirical results	40
2.4	Discussion and concluding remarks	47
B	APPENDIX	49
3		
	THE ULTIMATE CORPORATE OBJECTIVE IS SURVIVAL	57
3.1	Introduction	58
3.2	Model review	60
3.3	Data	63
3.4	Results	67
3.4.1	Parameter estimation	67
3.4.2	Do firm idiosyncrasies matter for profitability?	68

3.4.3	Are shorter lived firms different?	76
3.5	Concluding remarks	79
C	APPENDIX	81
C.1	Literature review	81
C.2	Estimation	88
C.2.1	Likelihood ratio test	88
C.2.2	Numerical analyses on $m$ and $\sigma$ estimators	88
C.3	Robustness checks	91
C.4	Effect of lifespan on average profitability and its volatility	93
4		
	A STATISTICAL EQUILIBRIUM APPROACH TO FORECASTING CORPORATE PROFITABILITY	95
4.1	Introduction	96
4.2	Data and forecasting methodology	99
4.2.1	Forecast design	102
4.2.2	Forecast evaluation	103
4.3	Competing models	105
4.3.1	Statistical equilibrium model of competitive firms	106
4.3.2	Ornstein-Uhlenbeck process	111
4.3.3	AR(I)MA-type models	112
4.4	Results	114
4.5	Concluding remarks	117
D	APPENDIX	119
D.1	Descriptive statistics	119
D.2	Estimation results	120
D.3	Restricted models	127
	BIBLIOGRAPHY	129

GIBRAT'S LAW REDUX: THINK PROFITABILITY INSTEAD OF GROWTH

---

## ABSTRACT

We argue that firm profitability can be conveniently characterized by one and the same diffusion process for all firms, akin to Gibrat's seminal idea of a common stochastic description for the dynamics of firm growth. Here we show that the diffusion of profit rates applies across all surviving corporations, irrespective of their size or industry, which is not true for growth rates. The diffusion process is empirically consistent with both the cross-sectional distribution of profit rates *and* the individual time series of corporate profit rates, leading to the notion of a statistical equilibrium. This has stark and unexpected implications for our understanding of individual destinies, since the model and data suggest that idiosyncratic efforts have no impact on the aggregate distributional outcome. Instead, idiosyncratic efforts merely affect the individual persistence of abnormal profits from the system-wide norm. Put differently, corporations have to participate in the same game and can merely choose the speed at which they play it, but they cannot bend the rules.

### 1.1 INTRODUCTION

Gibrat's rule of proportionate effect is probably the first and most influential stochastic model seeking to provide a dynamic law for the evolution of firm destinies. It claims that firm growth is independent of size, and that each firm's growth rate depends on random shocks to its current size, conventionally understood as independent draws from a normal distribution with *identical* mean and variance across all firms. After many decades of empirical research on Gibrat's "law", however, the literature seems to be in broad agreement that the law cannot adequately describe the growth dynamics across firms, essentially because the average and volatility of the individual time series depend on a multitude of factors like firm size, age, life-cycle, or numerous sectoral specificities (see, for instance, the surveys by Santarelli et al., 2006; Sutton, 1997). Therefore, it seems fair to say that a comprehensive and generally accepted stochastic theory of firm growth does not exist today. The lack of empirical support for Gibrat's original idea is hardly astounding because the idea simply lacks economic intuition, or as Sutton (p. 42) puts it in his well-known survey, "there is no obvious rationale for positing any general relationship between a firm's size and its expected growth rate, nor is there any reason to expect the size distribution of firms to take any particular form for the general run of industries." Essentially, the basic assumption of a common growth rate distribution that is invariant across firms of different sizes or industries seems questionable both from an empirical and theoretical point of view.

The purpose of the present paper is to show empirically that in contrast to firm growth, the rate of profit does in fact obey a common dynamic law across firms in many different sectors and industries, and that the expected rate of profit and even its volatility are independent of size. Put differently, we adapt Gibrat's original approach of looking for a common law of motion for firm destinies, but shift our attention away from growth rates and instead towards profit rates. Like Gibrat, whose growth model was inspired by the (cross-sectional) size distribution of firms, we also start from the

observation of the (cross-sectional) distribution of profit rates and the question which stochastic process could lead to such an outcome. Contrary to Gibrat, however, we rely on a well known principle of economics, namely the tendency for competition to equalize profit rates. The idea that the rate of profit is independent of size and tends to be equalized across competing uses has been one of the earliest and most deeply rooted theories in the history of economic thought, dating back to classical economics. It contends that market economies are driven by the perpetual reallocation of capital in search of profit rate equalization, and represents one of the most widely accepted theories of capitalism (see, e.g., Foley, 2006).

So instead of staying on Gibrat's well-trodden path of random growth rates that are drawn from a common distribution, we propose a dynamic law for the evolution of profit rates that turns out to be a diffusion process consistent with the cross-sectional distribution of profit rates (see Alfarano et al., 2012, for details). The process incorporates the principle that competition tends to equalize profit rates, a principle that provides the economic intuition for the very existence of a common distribution of profit rates across firms, and is absent in the case of growth rates. The diffusion process will serve as the theoretical foundation of our current empirical investigation and relies on three parameters: a system-wide average rate of profit, a system-wide dispersion or volatility measure of profit rates, and an idiosyncratic noise factor that determines the persistence of abnormal profits for individual firms.

Compared to our previous investigation, the major novelty in this paper is that we investigate the actual time series properties of profit rates, and not merely their cross-sectional distribution. It turns out that the model's assumption of a common location and dispersion parameter across all firms is clearly reflected in the time series data, which strongly suggests that the diffusion process is a useful description of the time evolution of individual firm profitability. Closed-form solutions for the transient density and the autocorrelation function of the diffusion process enable us to estimate the idiosyncratic noise levels with maximum likelihood, and to compute the half-life of abnormal profits (or adjustment speed towards the norm) for each firm.

Our main finding is that the process provides a very reasonable description of the profit rate evolution for all firms in the sample, irrespective of size or sectoral characteristics. In addition, the autocorrelation function (ACF) of the diffusion process is also consistent with the observed ACFs. Hence the diffusion process is not only consistent with the cross-sectional distribution of profit rates, but also with the time evolution of individual destinies, a situation that is the very definition of a *statisti-*

*cal equilibrium*.<sup>1</sup> To the best of our knowledge, the dynamic law governing corporate profitability is so far the only economic process that appears to exhibit statistical equilibrium, suggesting that profit rates are a most fundamental characteristic of the macroeconomy. Our economic interpretation of the statistical equilibrium property is that all surviving corporations are subject to the same competitive pressures of capital reallocation, irrespective of their industry or particular line of business. They all face the same profitability benchmark and volatility, while their idiosyncratic efforts merely affect the persistence of abnormal profits. In other words, survivors have to participate in the same game and can merely choose the speed at which they play it, but they cannot change the fundamental rules of the game, which apparently prescribe a common location and dispersion parameter for corporate profitability.

Interestingly, the empirical densities of both profit and growth rates are well described by exponential power or *Subbotin* distributions (see, e.g., Alfarano and Milaković, 2008; Alfarano et al., 2013; Bottazzi and Secchi, 2006; Bottazzi et al., 2001, 2002; Stanley et al., 1996, for growth rate distributions), often approximating a special case of the Subbotin distribution known as the Laplace distribution. In our experience, the similarity in the functional form of the cross-sectional distribution of profit and growth rates frequently evokes claims that the two quantities essentially measure the same thing and can be used interchangeably. This is false. Notice, firstly, that the two differ in their dimensionality because growth rates are logarithmic differences in size over time, while profit rates are defined as the ratio of a flow (of economic income over time) relative to a stock (of capital that generates this income). Secondly, our comparison of the statistical properties of growth and profit rates in the following section will show that there are pronounced differences in the parameterizations of the respective cross-sectional Subbotin distributions. Our dataset suggests that the growth rates of long-lived US non-banking corporations are even more leptokurtic than the Laplace distribution, which has recently been observed for the French manufacturing sector as well (see Bottazzi et al., 2011). The rate of profit, on the other hand, turns out to be significantly closer to the Laplace distribution than the corresponding growth rates in our sample and, more importantly, the volatility of profit

---

<sup>1</sup> Originating within statistical mechanics, and applying more broadly across the natural sciences, the term statistical equilibrium refers to the state of a system in which the average quantities governing this state are independent of time. In our reading of the economic literature, however, there has been very little interest in the concept so far, and there is currently no coherent definition of statistical equilibrium in the field. Econometric textbooks (see, e.g., Mills and Markellos, 2008, p. 10–11) refer to the situation we observe in our data in terms of ergodicity and stationarity, mentioning statistical equilibrium merely in passing (if at all). Other authors, starting with Foley (1994), use the term statistical equilibrium to denote the cross-sectional distribution that achieves the largest informational entropy in the variable of interest, but cannot provide information on the dynamic law that governs the state of the system. We refer the interested reader to Garibaldi and Scalas (2010) for a technical treatment of the statistical equilibrium notion in physics and its various flavors in economics.



rates turns out to be independent of size, which is not true for growth rates.<sup>2</sup> Thirdly, there are also pronounced differences in the autocorrelation structures of growth and profit rates. Previous empirical studies of the autocorrelations in growth rates yield inconclusive results, with positive autocorrelations in firm growth rates reported by Chesher (1979) and Geroski et al. (1997) for the UK, and by Weiss (1998) for Austria, while Boeri and Cramer (1992) and Goddard et al. (2002) observe negative serial correlations in German and Japanese data. Other studies do not find any significant autocorrelations in firm growth rates (see, for instance, Almus and Nerlinger, 2000; Geroski and Mazzucato, 2002; Lotti et al., 2003), whereas Coad (2007) reports that smaller French manufacturing firms exhibit negatively correlated growth rates, while larger firms display positive autocorrelations. Our results regarding long-lived US corporations indicate that there are no statistically significant autocorrelations in firm growth rates. In contrast, we show that profit rates do exhibit significantly positive autocorrelations.

We are of course aware that several of our empirical findings, especially those pertaining to growth rates, are already present in the literature. The purpose of our explicit statistical comparison in the next section is therefore not to claim novelty where none exists, but rather to convince the reader that profit rates are economically more fundamental and quantitatively more convenient than growth rates when it comes to finding a general law of motion for the destiny of corporations. Given the statistical properties of profit rates, the third section presents a model for the dynamic law that governs corporate profitability and accounts for the data. Since the diffusion coefficient (or noise level) remains the only source of idiosyncrasies in the statistical equilibrium model, we investigate its properties in the fourth section, finding that large, diversified, or capital intensive corporations exhibit the smallest diffusion coefficients, and therefore the most persistent deviations from the systemic rate of profit. The final section discusses our findings and concludes.

## 1.2 DATA

The data for this study are taken from Thomson Reuters' *Datastream* and consist of annual observations for the sales, operating income, total assets, number of employees, and market value of publicly traded US companies. According to the database, a total of 6,860 firms have been present in the market for at least one year from 1980–2011, and have operated in at least one of the 78 different sectors listed in Table 2 in Appendix A.1. Unlike many previous studies that typically focus on the manufacturing sectors (SIC codes 20 to 39), our present analysis considers a diverse set of firms

---

<sup>2</sup> The classical claim that profitability should be independent of size has already been established in the empirical literature (see, e.g., Alexander, 1949; Amato and Wilder, 1985; Goddard et al., 2005; Hall and Weiss, 1967; Marcus, 1969; Whittington, 1980), and we also find that the profitability of surviving US corporations is independent of size in our sample. The important point here concerns the volatility in profit and growth rates. In our sample, the rate of corporate growth would also seem independent of size but its volatility, unlike the volatility of profit rates, is not.

across the different sectors, and merely excludes banks (SIC codes 60 and 61) because their balance sheets exceed those in other sectors by at least an order of magnitude.

We focus on long-lived or “surviving” firms that we define as companies operating in the market for the entire time span from 1980 to 2011. This panel contains 522 companies that account on average for more than seventy percent of market capitalization, total assets and employment in the sample. The importance of such a ‘granular’ view of the economy has recently been forcefully argued by Gabaix (2011), who finds that about one third of variations in US GDP growth can be attributed to the idiosyncratic destinies of the largest one hundred corporations. This is particularly true for power-law distributed firm sizes (see Axtell, 2001), where the largest firms contribute disproportionately to aggregate output, and the empirical size distribution of companies in our sample also indicates a power-law tail

$$f(S) = \alpha S_{\min}^{\alpha} S^{-(\alpha+1)} \quad (1)$$

for  $S \geq S_{\min}$ , with a tail index  $\alpha$  close to unity, also known as Zipf’s law.<sup>3</sup> The power law tail contains between 149 (in 1980) and 214 (in 2011) companies, and on average long-lived corporations account for seventy percent of the companies in the upper tail. Hence the impact of these corporations on the economy is non-negligible, and we consider their dynamics as a crucial determinant of macroeconomic fluctuations.

For each company, we compute annual (logarithmic) growth rates  $g$  of firm size  $S$ ,

$$g_{i,t} = \log(S_{i,t+1}) - \log(S_{i,t}), \quad (2)$$

where  $i$  runs over firms and  $t$  denotes time. We consider sales, total assets, number of employees, and market value as proxies for firm size.<sup>4</sup> Our proxy for the profit rate  $p$  is the *return on assets*,

$$p_{i,t} = \frac{I_{i,t}}{A_{i,t}}, \quad (3)$$

where  $I$  denotes operating income, and  $A$  denotes total assets.

<sup>3</sup> The critical cut-off value  $S_{\min}$  was identified with a Hill plot, which graphs the estimated tail index as a function of tail size. For every year we observe a plateau around the benchmark  $\alpha = 1$  after a few initial oscillations, suggesting that the Zipf law is a fairly robust feature of the size distribution’s upper tail. Concerns regarding a potential ‘survivorship bias’ are addressed in the final section.

<sup>4</sup> We will mostly report data for growth in total assets since both referees suggested to focus on a single growth quantity for better legibility. Results for the other quantities are qualitatively in line with results for growth in total assets. As one might expect, the growth rate of market value turns out to be the most volatile quantity; on average, its standard deviation exceeds the standard deviation of profit rates by a factor greater than three. Employment growth appears more volatile than growth in sales or total assets, but the latter are still approximately twice as volatile as profit rates. This material is available upon request.

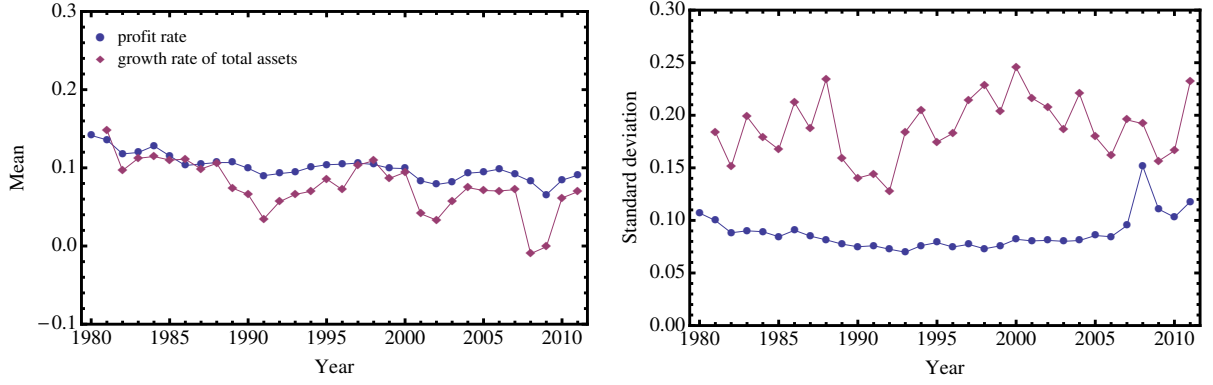


Figure 1: Time evolution of the mean and standard deviation of corporate profit and growth rates. Lines between data points are visual aids.

### 1.2.1 Empirical densities

To understand the distributional properties of growth and profit rates, it is instructive to first consider the time evolution of their means and standard deviations in Figure 1. Notice that the mean and standard deviation of profit rates are remarkably stable compared to their growth rate counterparts. The stability of the average profit rate becomes most apparent when we look at the dot-com bubble and the recent financial crisis. During those years a massive drop in market demand was reflected in sizable decreases in firm growth and sometimes even in firm size. The adverse effects on firm profitability, however, appear very moderate in comparison to the growth rate series.<sup>5</sup> Although Figure 1 merely shows the average behavior, we do observe a decline in growth rates that is more pronounced than the decline in profitability.<sup>6</sup> Yet the first two moments still provide less information than the distributions of growth and profit rates, which we turn to next.

In the recent literature on growth rate distributions, it is common practice to eliminate possible trends in firm size by considering the normalized (logarithmic) size

$$s_{i,t} = \log(S_{i,t}) - N^{-1} \sum_{i=1}^N \log(S_{i,t}), \quad (4)$$

<sup>5</sup> One could speculate that adverse demand shocks induce firms to reduce costs (number of employees) or the scope of their operations (total assets), thereby mitigating the effects of decreasing sales on profitability.

<sup>6</sup> This non-trivial stability of the profit rate over time has also been pointed out by Mundt et al. (2014), who analyze data from more than 30,000 publicly traded firms in more than forty countries that account for about ninety percent of world GDP. Therefore we would like to think that our present findings do not just reflect a peculiarity of the US data.

which is obtained by subtracting the average (log) size of all long-lived firms from the (log) size of company  $i$ . Then the normalized growth rate is defined as the first difference of (4)

$$\tilde{g}_{i,t} = s_{i,t+1} - s_{i,t}. \quad (5)$$

Profit rates, on the other hand, are not normalized in any way and simply remain in the raw form (3). In order to fit the empirical distributions of growth and profit rates in our sample, we follow standard procedure in the field and employ the *exponential power distribution* first suggested by Subbotin (1923),

$$f(x) = \frac{1}{2\sigma\alpha^{1/\alpha}\Gamma(1+1/\alpha)} \exp\left(-\frac{1}{\alpha} \left|\frac{x-m}{\sigma}\right|^\alpha\right), \quad (6)$$

where  $\alpha, \sigma \in \mathbb{R}^+, m \in \mathbb{R}$ , and  $\Gamma(\cdot)$  denotes the Gamma function. The distribution is characterized by three parameters: a location parameter  $m$ , a scale parameter  $\sigma$ , and a shape parameter  $\alpha$  that is responsible for qualitative differences in the distribution, in particular its kurtosis. It is readily verified that the Subbotin density includes the Laplacian ( $\alpha = 1$ ) and the Gaussian ( $\alpha = 2$ ) as special cases.

Figure 2 presents the pooled empirical densities of profit rates and normalized growth rates, as well as the corresponding Subbotin fit obtained from maximum likelihood estimation of the parameters, reported in Table 3 (see Appendix A.2). The parameter estimates of the pooled empirical distribution of profit rates are denoted by  $\hat{\alpha}$ ,  $\hat{\sigma}$  and  $\hat{m}$ . We find that the empirical densities of both profit and growth rates are clearly non-Gaussian, and can be reasonably well approximated by a symmetric Subbotin distribution. The empirical density of profit rates exhibits a “linear tent-shape” on semi-log scale that is characteristic of the Laplace distribution, while the growth rate distribution is more leptokurtic than the Laplace.<sup>7</sup> This visual impression is confirmed by maximum likelihood estimates of the shape parameter  $\alpha$ , which are significantly smaller than unity for growth rates, while the maximum likelihood estimates of the scale parameter  $\sigma$  illustrate that growth rates are also significantly more volatile than profit rates.

To check whether our results are affected by the aggregation of data from different years, we have also estimated  $\alpha$  and  $\sigma$  for every single year during the period 1980-2011. As Figure 3 illustrates, there is a remarkable year-to-year stability of the Laplace distribution for profit rates, with rather small fluctuations in the parameter values over time. In 25 out of 32 years the estimated shape parameter is consistent with a Laplace distribution at the 95% confidence level. Since maximum likelihood estimates of the shape parameter are quite sensitive to outliers, we investigated the relatively small values of the shape parameter in the last four years, and it turns out that they are in

<sup>7</sup> Some studies (see, for instance Amaral et al., 1997; Bottazzi et al., 2011) investigate the empirical distribution of “rescaled” growth rates that are divided by their standard deviations conditional on firm size. To a certain extent, this procedure brings the parameter estimates closer to a Laplace fit, yet we obtain a significantly better Laplace fit for profit rates using merely raw data on the ratio of operating income to total assets in our dataset.

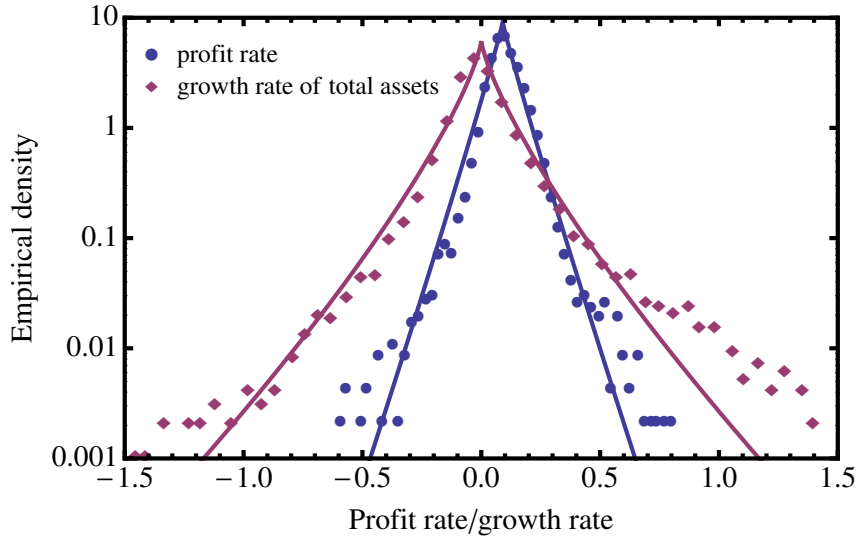


Figure 2: Pooled empirical densities of annual profit and growth rates for long-lived (non-banking) corporations in the United States during the period 1980-2011. The mode of the profit rate distribution is  $\hat{m} = 9.3\%$ . Due to the normalization of growth rates, their distribution is centered around zero. The solid curves show the Subbotin fit obtained by maximum likelihood estimation of the shape and scale parameter (standard errors in parentheses), yielding estimates of  $\hat{\alpha} = 0.95$  (0.01) and  $\hat{\sigma} = 0.0570$  (0.0005) for profit rates, and  $\hat{\alpha} = 0.76$  (0.01) and  $\hat{\sigma} = 0.0977$  (0.0010) for growth rates.

fact due to very few extreme observations. Eliminating, for instance, merely the two most extreme profit rates at both sides of the spectrum leads to estimates for  $\alpha$  that cannot be distinguished from the Laplace benchmark ( $\alpha = 1$ ) at the 95% confidence level. As far as growth rates are concerned, we can reject the null hypothesis  $\alpha = 1$  in approximately 90% of cases. Thus we continue under the assumption that the Laplace distribution is a reasonable benchmark for the distribution of profit rates, while growth rates are more leptokurtic.

### 1.2.2 Autocorrelations

Inspection of the line charts for a dozen randomly chosen time series of growth and profit rates indicates that profit rates are substantially more persistent than growth rates. To properly quantify this graphical impression, we consider their autocorrelation function (ACF). For the estimation of the autocovariance function, however, we must consider that the number of observations per time series is quite small in our sample. In case of an autocorrelated process and a small number of observations, subtracting the sample mean from the observations in the autocovariance function leads to a systematic underestimation of the true autocorrelation (see, for instance,

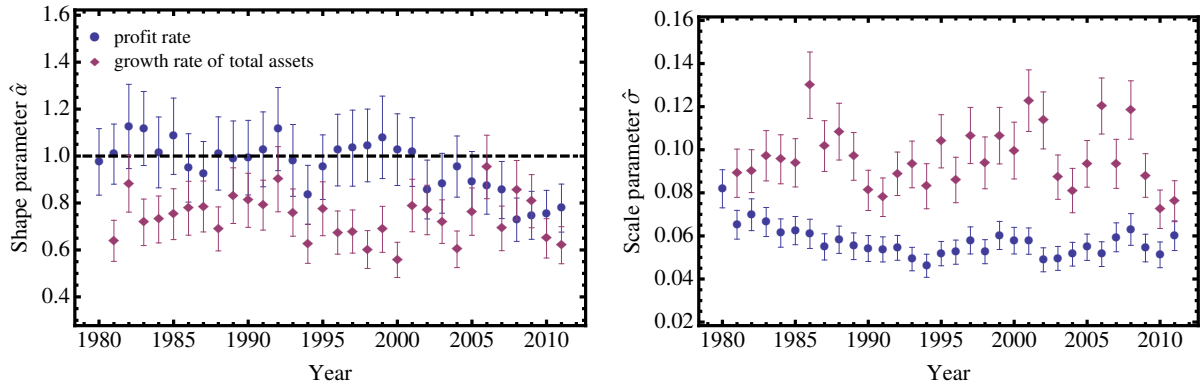


Figure 3: Year-by-year maximum likelihood estimates of the Subbotin shape (left panel) and scale (right panel) parameters. Error bars show two standard errors.

Fuller, 1996). Intuitively, this negative bias stems from the fact that the autocorrelation coefficient is a scaled sum of cross-products of deviations of  $X_t$  from its mean. For each time series these deviations must sum to zero by construction, so that negative deviations must eventually be followed by positive deviations on average and vice versa. Therefore, the expected value of cross-products of deviations is negative (see Campbell et al., 1996, p. 46). In order to mitigate this negative bias, we replace the estimated mean of each individual time series with the median  $\hat{m}$  of the pooled empirical density in the autocovariance function.<sup>8</sup>

Figure 4 presents a box-and-whisker plot for the 522 estimated autocorrelation functions of the profit rate and growth rate time series. Our analysis suggests that statistically significant autocorrelations in growth rates can only be found in relatively few time series: serial correlation seems to be completely absent in annual growth rates of market value (in line with the weak-form efficient market hypothesis of Fama, 1991), which is consistent with many previous findings in the pertinent empirical literature (see, for instance, Cont, 2001, for a review of the stylized facts of financial returns). The annual growth rates of sales, total assets, and employment appear to be slightly more persistent, yet the estimated autocorrelation coefficients cannot be distinguished from zero at the 95% confidence level in the vast majority of cases. If present at all, we find that autocorrelation in growth rates is very weak, consistent with previous results by Bottazzi et al. (2001) and Bottazzi and Secchi (2005). These findings can be interpreted as evidence against the “optimal size” hypothesis since one should observe pronounced positive autocorrelations in growth rates as firms approach some optimal size.<sup>9</sup>

<sup>8</sup> The reason why this particular substitution is preferable will be clarified in Section 1.2.3, and is essentially based on the statistical equilibrium property of the profit rate series. The relevance of the negative bias for our subsequent diffusion model is illustrated in Figure 16 (see Appendix A.4).

<sup>9</sup> This is particularly true in case of (non-linear) adjustment costs that prohibit firms from instantly attaining their optimal size. Instead, these firms would grow gradually by equating the marginal gains from having a larger size with the marginal costs of growing (see Coad, 2007).

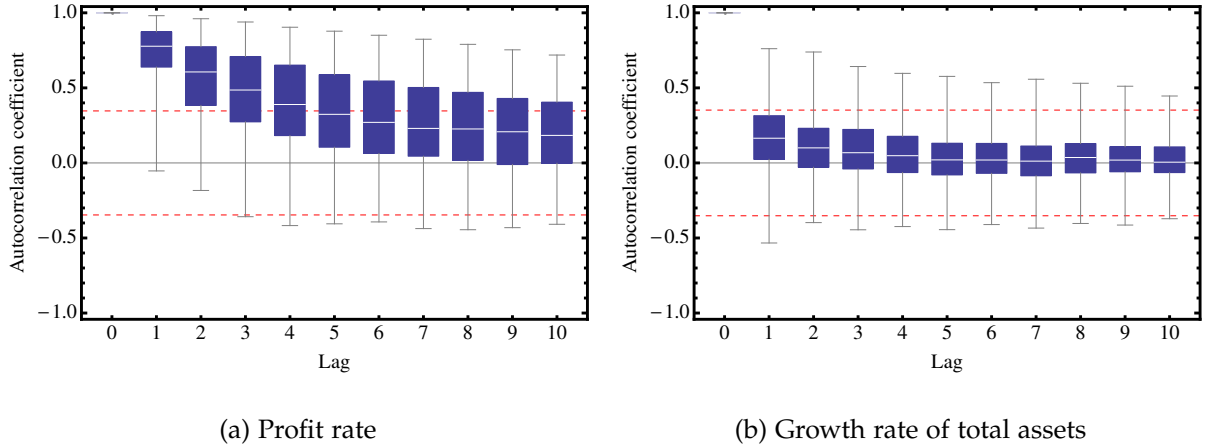


Figure 4: Box-and-whisker plot for the estimated autocorrelation functions of profit rates and growth rates. The boxes include the 25 percent quantile, the median, and the 75 percent quantile. The red dashed lines show the 95 percent confidence interval under the null hypothesis of zero autocorrelations. The interval has been computed as  $\pm 1.96/\sqrt{T}$ , where  $T = 32$  is the length of the profit rate time series.

Profit rates, on the other hand, exhibit strong positive autocorrelations. Similar results have been reported in the so-called *persistence of profits* literature, which finds significantly positive first-order autoregressive coefficients in time-series regressions of profit rates (for a recent take on the subject see, for instance, Cable and Mueller, 2008). Notice, however, that these models typically approach the dynamics of firm profitability via stationary AR(1) processes, and hence are misspecified because their stationary distribution is Gaussian, yet the previous subsection shows that empirical profit rate distributions are much closer to the Laplace. Since we cannot rule out a negative bias in the estimated autocorrelation coefficients of profit rates, it seems imprudent to specify the number of statistically significant time lags for the non-parametric analysis in Figure 4. Instead we will introduce the correlation time of our subsequent diffusion model as an alternative measure of profit persistence in section 1.4.1.

### 1.2.3 Size (in)dependence

The law of proportionate effect is conventionally understood as a multiplicative stochastic process whereby a firm's current size is the result of a sequence of independent and identically distributed growth shocks. According to the central limit theorem, the growth rate distribution should then be Gaussian, and the corresponding firm size distribution should be log-normal.<sup>10</sup> While the hypothesis of proportionate random

<sup>10</sup> From a technical point of view, one should be more precise and additionally state that the length of the sequence of growth shocks (and the mean and variance of the resulting distribution) is the same for all firms.

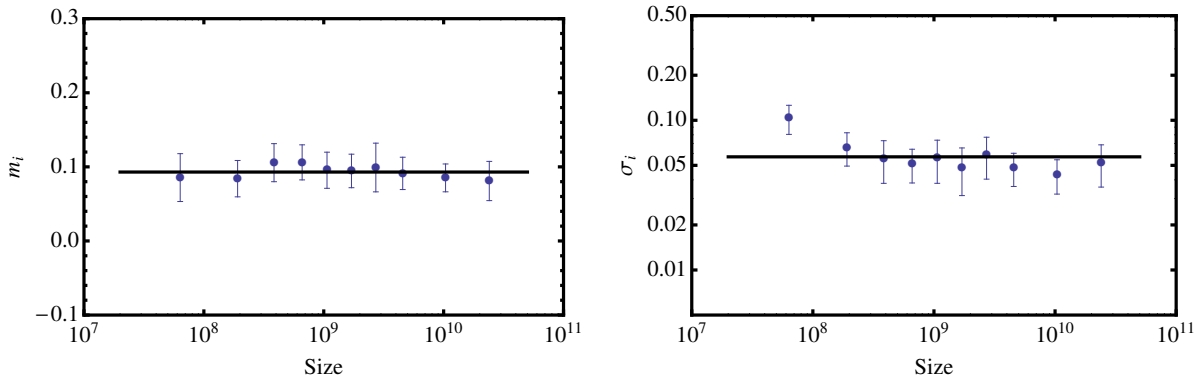


Figure 5: Location ( $m_i$ ) and dispersion ( $\sigma_i$ ) of firm profit rates as a function of firm size. Points represent binned data and have been computed in the following way: for each time series,  $i = 1, \dots, 522$ , we calculate a firm's median profit rate  $m_i$  and corresponding mean absolute deviation  $\sigma_i$ . Then we split the firms according to their median size into ten (almost) equipopulated bins. The points represent the average  $m_i$  and  $\sigma_i$  of the approximately 52 firms in each bin. The black lines represent the phenomenological values  $\hat{m} = 9.3\%$  and  $\hat{\sigma} = 5.7\%$  of the empirical profit rate distribution. Error bars correspond to one median absolute deviation.

growth is useful to explain the considerable heterogeneity in firm size, it lacks an economic justification as already pointed out in the introduction.

There is, on the other hand, good reason for profit rates to be independent of size. Profit rates are central to economic competition since they guide the allocation of capital across competing uses in different sectors and industries. Capital seeks out abnormally profitable activities independent of their size, because it is the rate of return to invested capital (say, ten percent), and not the absolute return (say, ten million currency units) that guides the allocation of capital. In the absence of further information, one should therefore expect both the location parameter  $m$  of the profit rate distribution, and the dispersion parameter  $\sigma$  to be independent of firm size. In order to judge how well the data reflect this prediction, we consider the median and mean absolute deviation as the location and dispersion measures, because they correspond to the maximum likelihood estimators of  $m$  and  $\sigma$  when sampling from a Laplace distribution (see, for instance, Johnson et al., 1995; Kotz et al., 2001). As illustrated in Figure 2, the Laplace is a reasonable benchmark for the pooled profit rate distribution, and hence we denote the parameter estimates from the pooled cross-sectional distribution by  $\hat{m}$  and  $\hat{\sigma}$ .

To further fix notation, let  $m_i$  and  $\sigma_i$  denote the median and mean absolute deviation of the profit rate time series of firm  $i$ . Figure 5 suggests that both the median and mean absolute deviation of profit rates are rather homogeneous across different size classes, and are reasonably close to the unconditional values  $\hat{m} = 0.093$  and  $\hat{\sigma} = 0.057$  of the pooled empirical profit rate distribution. We cannot rule out the existence of a small negative bias for the smallest size bin, yet this bias is caused by around 15



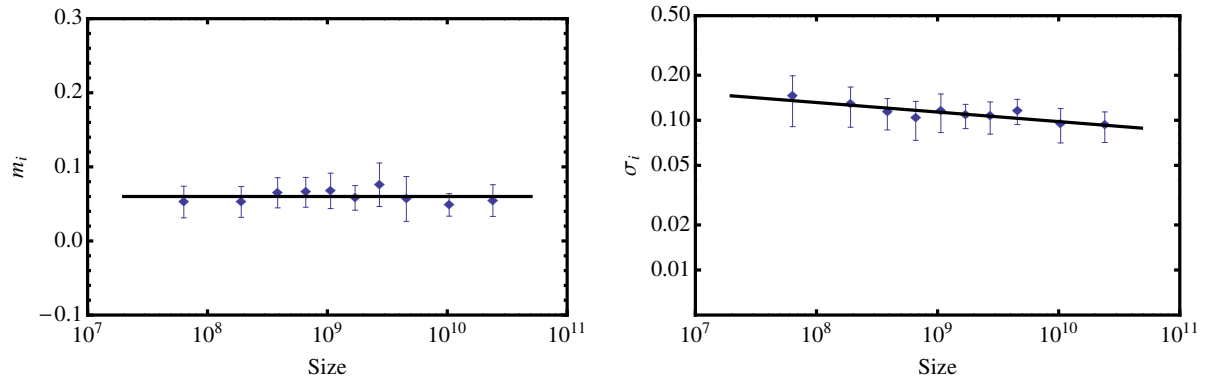


Figure 6: Location ( $m_i$ ) and dispersion ( $\sigma_i$ ) of growth rates as a function of firm size. The binning procedure is the same as in Figure 5. A least squares power law fit for the relation between the dispersion of growth in total assets and size yields  $-0.06 \pm 0.01$ . The use of alternative size measures leads to the estimates  $-0.10 \pm 0.02$  for sales,  $-0.10 \pm 0.02$  for the number of employees, and  $-0.09 \pm 0.02$  for market value (not shown here). Error bars correspond to one median absolute deviation.

to 25 out of 522 firms, that is by about 3 to 5 percent of all long-lived companies.<sup>11</sup> The visual impression that  $m_i$  and  $\sigma_i$  are virtually the same for the vast majority of firms is supported by linear regressions, which yield slope coefficients that cannot be distinguished from zero at the usual confidence levels once the smallest size bin is excluded from the analysis (see Tables 4 and 5 in Appendix A.5 for details). The intercept in the linear regressions provides rather limited information since firm sizes span several orders of magnitude and are very large to begin with, so that an extrapolation of any linear relationship to size zero is hardly meaningful. In any case, the values of  $m_i$  and  $\sigma_i$  in each bin cannot be distinguished from  $\hat{m}$  and  $\hat{\sigma}$  at the usual significance levels. The remarkable similarity between the parameters of the pooled cross-sectional distribution and the individual corporate time series strongly suggests that profit rates are generated by a common dynamic law.<sup>12</sup>

Figure 6 repeats the analysis for the growth rates of firm size. While the location parameter of growth rates is not markedly affected by size, we find a clear inverse relation between the dispersion of growth rates and company size, in line with previous studies that report power-laws with scaling exponents close to  $-0.15$  (see, for instance, Amaral et al., 1997; Bottazzi and Secchi, 2003; Stanley et al., 1996). Fitting a power law to our data yields scaling exponents ranging from  $-0.06 \pm 0.01$  for growth in total assets to  $-0.10 \pm 0.02$  for sales and employment growth rates. In contrast to profit rates, removing the smallest size bin from the growth rate analysis does not lead to

<sup>11</sup> Twelve (thirteen) of the twenty-five corporations with the lowest  $m_i$  (highest  $\sigma_i$ ) operate in just four industries with SIC codes 13, 36, 38, and 67.

<sup>12</sup> One referee suggested to check the validity of our findings by repeating the analysis for the subset of manufacturing firms (SIC codes 20 to 39). It turns out that all our central results (like the distributional regularities, autocorrelation structures, and size independence of profit rates) carry over to this subsample (see Appendix A.3 for details).

significantly different slope coefficients,<sup>13</sup> and preserves the scaling of the dispersion of growth rates with firm size, which the pertinent literature typically ascribes to firm diversification.

### 1.3 MODEL

The preceding analysis suggests that profit rates are characterized by a stationary cross-sectional distribution, and that firm-level time series exhibit persistent autocorrelations. In addition, the location and dispersion of the individual series are independent of size and very close to the location and dispersion estimates  $\hat{m}$  and  $\hat{\sigma}$  of the pooled cross-sectional distribution. These properties of corporate profit rates establish a major difference to growth rates, and would seem to represent a more immediate way to study the competitive behavior of corporations, at the very least from a statistical point of view. Inspired by the empirical densities of cross-sectional profit rates, Alfarano et al. (2012) have recently introduced a diffusion process with a stationary Laplace distribution. We argue here that the process is not only consistent with the observed cross-sectional distribution, but also with the time series properties of surviving corporations, including their autocorrelation structures. To the best of our knowledge, the Laplace diffusion introduced below is the only model for the dynamics of firm profitability that is consistent with both the cross-sectional distribution and the individual time series properties of profit rates.

#### 1.3.1 Diffusion

Alfarano et al. (2012) propose the stochastic differential equation

$$dX_t = -\frac{D}{2\sigma} \text{sign}(X_t - m) dt + \sqrt{D} dW_t, \quad (7)$$

to model the dynamic evolution of firm profitability, where  $X_t$  denotes the profit rate,  $\sigma$  is a dispersion parameter,  $\text{sign}(\cdot)$  denotes the signum function,  $m$  is the average rate of profit, and  $dW_t$  are Wiener increments. The (square root of the) constant term  $D$  determines the noise level in the (random) second term, but notice that it also influences the strength of the reversion to  $m$  in the (deterministic) first term. From an economic point of view, this part of the stochastic process reflects the negative feedback mechanism of classical competition: capital will seek out sectors or industries where the profit rate is higher than the economy-wide average, typically attracting labor, raising output, and reducing prices and profit rates in the sector. This provides an incentive for capital to leave the sector, leading in turn to higher prices and profit rates for the surviving firms. Notice that the first term (or *drift function*) of the stochastic model (7) does not depend on the current profitability of a firm. The second term

---

<sup>13</sup> To be precise, the cited papers consider the relationship between size and the standard deviation of growth rates. We have also estimated the power law exponents for this relationship and find that the results are very similar to the ones we report for the mean absolute deviation.

(or *diffusion function*) is governed by random Wiener increments, which incorporate all idiosyncratic factors affecting firm profitability. In the particular case (7), the noise level  $\sqrt{D}$  is constant over time and independent of the profit rate. Another particular feature of the diffusion (7) is that the drift and the diffusion function are intertwined, since the variance of the idiosyncratic noise term affects the speed of adjustment towards the system-wide average. It can be interpreted in the sense that competition simultaneously generates fluctuations in corporate destinies and convergence to the average rate of profit.

The notion of statistical equilibrium rests on the idea that all surviving corporations are subject to the same stochastic process (7), with common parameters  $m$  and  $\sigma$  that match the location and dispersion parameters of the stationary distribution of the process,

$$f_S(x; m, \sigma) = \frac{1}{2\sigma} \exp\left(-\left|\frac{x - m}{\sigma}\right|\right). \quad (8)$$

It is easily verified that (8) is obtained from (6) for  $\alpha = 1$ . The empirical analysis in the preceding section therefore suggests that  $m$  and  $\sigma$  can be readily observed from the pooled empirical density of profit rates, and would apply to the individual destinies of all surviving corporations, regardless of their size or industry. In other words, statistical equilibrium describes a situation where the profit rate of each surviving corporation reverts to the same *systemic* rate of profit, and fluctuates around it with the same (systemic) variability. That means that the only source of firm-specific effects in the model originates from the diffusion coefficient  $D$ , because the unconditional equilibrium distribution (8) does not depend on this parameter. Accordingly, the statistical equilibrium model leaves a single degree of freedom for idiosyncrasies in corporate profitability.

### 1.3.2 *Transient density*

A useful alternative representation of a diffusion process is provided by its transient density (or Fokker-Planck equation, see Risken, 1996), which describes the time evolution of the stochastic system by means of a second-order partial differential equation,

$$\frac{\partial p(x, t)}{\partial t} = -\frac{\partial}{\partial x}(A(x; D)p(x, t)) + \frac{1}{2} \frac{\partial^2}{\partial x^2}(B(x; D)p(x, t)), \quad (9)$$

where  $A(x; D)$  and  $B(x; D)$  are the drift and diffusion functions of the underlying diffusion process, and

$$p(x, t) = f(x, t|x_0, t_0) \quad (10)$$

denotes the conditional probability density for a transition from state  $x_0$  at time  $t_0 = 0$  to state  $x$  at time  $t$ . For the particular diffusion (7) with zero mean, that is

$$Z_t = X_t - m, \quad (11)$$

and initial condition  $f(z, 0|z_0, 0) = \delta(z - z_0)$ , where  $\delta(\cdot)$  denotes Dirac's delta function, Toda (2012) demonstrates that a closed-form solution to (9) exists under appropriate boundary conditions and is given by

$$f(z, t|z_0, 0) = \frac{1}{\sqrt{2D\pi t}} \cdot \exp\left(-\frac{(z - z_0)^2}{2Dt} - \frac{1}{2\sigma}(|z| - |z_0|) - \frac{D}{8\sigma^2}t\right) + \frac{1}{2\sigma} \exp\left(-\frac{1}{\sigma}|z|\right) \Phi\left(-\frac{|z| + |z_0| - (Dt)/(2\sigma)}{\sqrt{Dt}}\right), \quad (12)$$

where  $\Phi(\cdot)$  denotes the cumulative distribution function of the standard normal. The closed-form solution of the Fokker-Planck equation in (12) allows us to estimate the idiosyncratic diffusion coefficient by maximum likelihood, and it is also helpful in finding a closed-form solution for the autocorrelation function of the diffusion process (7).

### 1.3.3 Autocorrelation function

For stationary Markov processes, the autocorrelation function obeys the textbook formula (see, for instance, Kampen, 1992)

$$\kappa(\tau) = \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} dz_0 z z_0 f(z, \tau|z_0, 0) f_S(z_0), \quad (13)$$

where  $f_S$  denotes the stationary density. Here the stationary density corresponds to (the zero-mean-shifted version of) equation (8), and the transient density  $f$  obeys (12). In this case, Touchette et al. (2010) show that the autocorrelation function of (7) is characterized by an (asymptotic) exponential decay,<sup>14</sup>

$$\kappa(\tau) = \frac{1}{6\sqrt{\frac{2\pi D\tau}{\sigma^2}}} \exp\left(-\frac{D\tau}{8\sigma^2}\right) \left\{ \left( \frac{\sqrt{\frac{\pi D\tau}{2\sigma^2}}}{2} \exp\left(\frac{D\tau}{8\sigma^2}\right) \operatorname{erfc}\left(\frac{\sqrt{\frac{D\tau}{\sigma^2}}}{2\sqrt{2}}\right) - 1 \right) \left( \frac{D^3\tau^3}{8\sigma^6} + \frac{3D^2\tau^2}{2\sigma^4} - \frac{6D\tau}{\sigma^2} + 24 \right) + \frac{D^2\tau^2}{2\sigma^4} + 24 \right\}. \quad (14)$$

Figure 15 (see Appendix A.4) illustrates the probability density and autocorrelation function of simulated realizations of the diffusion process in equation (7). While the model is consistent with the distributional and autocorrelation properties of empirical profit rates, the good fit between the estimated and theoretical autocorrelation func-

<sup>14</sup> The pre-factor stems from the non-linear nature of the drift function in (7).

tion only occurs for time series that are sufficiently long. For shorter time series, we do observe a negative bias in the estimated autocorrelation function. This, however, implies that the persistence of abnormal profits is actually even stronger than Figure 4 suggests. Given (7), we can avoid the negative bias in the persistence of abnormal profits by using the theoretical autocorrelation function. Estimating the persistence of abnormal profits then boils down to estimating the diffusion coefficient  $D$  from the transient density (12).

#### 1.4 RESULTS

In the diffusion model (7), the persistence of profits is determined by the drift function, hence the speed of convergence towards the systemic rate of profit depends on two parameters: the diffusion coefficient  $D$  and the scale parameter  $\sigma$ . In statistical equilibrium all firms are subject to the same location and scale parameters  $m$  and  $\sigma$ , so the diffusion coefficient  $D$  remains as the only source of idiosyncratic differences in the profitability of surviving corporations. If  $\sigma$  is the same for all corporations, then the noise level  $D_i$  measures the persistence of abnormal profits directly, and can be interpreted in the sense that firms with larger diffusion coefficients are prone to larger shocks in their profitability, while their abnormal profits do not persist for long. Conversely, firms with smaller diffusion coefficients are on average subject to smaller shocks, while their abnormal profits are more persistent. In order to estimate the diffusion coefficient for each profit rate series, we apply the maximum likelihood method to the solution (12) of the Fokker-Planck equation.

##### 1.4.1 Estimation of the diffusion coefficient

Given discrete annual observations, we estimate the diffusion coefficients for each firm by numerically minimizing the negative log-likelihood

$$-\log \mathcal{L}(D_i) = -\log f_S(z_{i,0}) - \sum_{t=0}^{T-1} \log f(z_{i,t+1}|z_{i,t}; D_i) \quad (15)$$

with respect to  $D_i$ , where  $f_S(z_{i,0})$  is the stationary Laplace density of some initial state  $z_{i,0}$ , and  $f(z_{i,t+1}|z_{i,t}; D_i)$  is the solution of the transient density (12) evaluated for each observation  $z_{i,t+1} = p_{i,t+1} - \hat{m}$  at time  $t + 1$  conditional on the previous observation  $z_{i,t}$  at time  $t$ . Equipped with the estimated coefficients, we then compute the *speed of adjustment* (or *characteristic time scale* or *relaxation time*) of the profit diffusion from (14) as the number of years that are necessary for the autocorrelation function to reach the value one half.<sup>15</sup>

<sup>15</sup> We consider the half-life definition in order to account for the non-linear nature of the diffusion process. The usual choices that are typically based on the dominant exponential term would in fact neglect such effects. Essentially we are solving the implicit equation  $\kappa(\tau_i; D_i, \hat{\sigma}, \hat{m}) = 0.5$  for  $\tau_i$ .

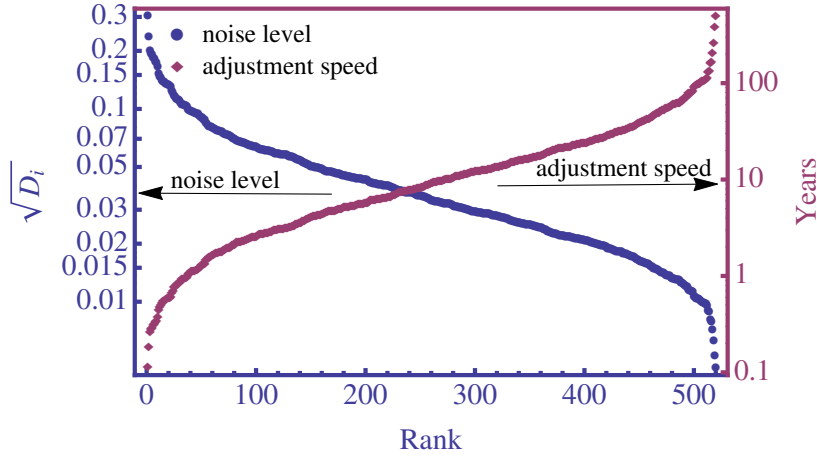


Figure 7: Sorted estimates of the noise level  $\sqrt{D}$  (left axis) and corresponding relaxation time of abnormal profits (right axis) for long-lived US corporations. The latter shows the number of years that are necessary for the autocorrelation function (14) to reach the value  $1/2$ . The arrows indicate the median noise level of around 3.3% p.a., and the corresponding median half-life of abnormal profits (around 9 years).

Figure 7 presents the estimated diffusion coefficients and corresponding half-life of abnormal profits, where we observe a pronounced variability in the diffusion coefficients that translates into heterogeneous time horizons for the dissipation of abnormal profits. The median diffusion coefficient is  $D_{\text{med}} \approx 1.1 \times 10^{-3}$ , implying a standard deviation of the idiosyncratic noise in the diffusion equation of  $\sqrt{D_{\text{med}}} \approx 3.3\%$  per annum, which corresponds to a longitudinal relaxation time of about nine years. For some firms, however, the diffusion coefficients are very small and imply relaxation times that are much longer than the length of the observed time series. Analyzing these firms in more detail, we find that considerable fractions are made up by utilities or insurance companies, with total assets far above the sample average, high capital intensity, and relatively steady profit series in comparison to other sectors. At this point we can merely speculate that entry and exit barriers in these sectors, probably stemming from large capital requirements, prevent a smooth and frictionless reallocation of capital in search of profit rate equalization.

Firms with relatively short relaxation times, on the other hand, disproportionately often operate in business sectors with SIC code 36 (electronical equipment), 38 (measuring instruments), and 13 (oil and gas extraction). Intuitively, the latter is charac-

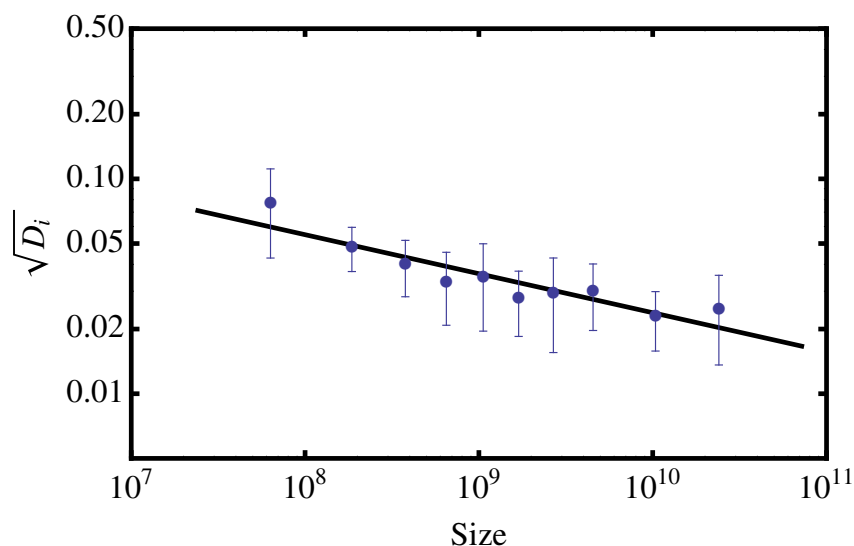


Figure 8: Estimates of the noise level ( $\sqrt{D}$ ) vs firm size on double-logarithmic scale. A least-squares power-law fit yields a scaling exponent of  $0.18 \pm 0.03$  for total assets as a measure of size. Using alternative size measures yields the estimates  $0.17 \pm 0.03$  for sales,  $0.16 \pm 0.02$  for the number of employees, and  $0.16 \pm 0.03$  for market value. Error bars correspond to one median absolute deviation.

terized by a high degree of uncertainty, while large changes in profitability for the former two sectors might be caused by operating leverage effects.<sup>16</sup>

#### 1.4.2 What determines the diffusion coefficient?

The observed heterogeneity in the diffusion coefficient raises the question whether firm or industry characteristics affect the persistence of abnormal profits. We will focus on what are perhaps merely the most obvious attributes, and consider here the impact of size, diversification, and capital intensity on the persistence of abnormal profits.

#### 1.4.3 Firm size

While the data suggest that size basically does not influence the rate of profit, we can ask whether size instead has an impact on the diffusion coefficient? On average larger corporations appear more stable and are affected by smaller idiosyncratic shocks to their profitability than smaller entities. The double-logarithmic plot in Figure 8 sug-

<sup>16</sup> Operating leverage increases with the proportion of fixed in relation to variable operating costs. During demand surges, high operating leverage could well lead to larger profits, but it also makes firms more vulnerable as they cannot readily cut expenses to absorb plummeting demand when most costs are tied up in machinery, plants, real estate, or distribution networks.

gests that the standard deviation of the idiosyncratic noise scales with size according to a power law

$$\sqrt{D} \sim \alpha S^{-\beta}. \quad (16)$$

To avoid distortions arising from booms and busts in single years, we calculated the mean size of each surviving corporation during the period 1980-2011,<sup>17</sup> and divided the sorted values into deciles, calculating the median size and  $\sqrt{D_i}$  in each decile. Fitting a power relation

$$\log \sqrt{D} = \log \alpha - \beta \cdot s \quad (17)$$

to the data yields least squares estimates of  $\beta = 0.18 \pm 0.03$  for total assets,  $\beta = 0.17 \pm 0.03$  for sales,  $\beta = 0.16 \pm 0.02$  for the number of employees, and  $\beta = 0.16 \pm 0.03$  for market value. All estimates are significant at the one percent level and indicate an inverse relationship between size and the noise level, so the larger a corporation the more persistent its abnormal profits tend to be.

#### 1.4.4 *Diversification*

In order to proxy the degree of corporate diversification, we consider Datastream's product segment decomposition of corporate revenues. The data associate segment-level SIC codes with the corresponding revenues of each company, and we use the product segment data to compute three common measures of corporate diversification: segment count, Herfindahl index, and entropy.

The first measure literally counts the number of sectors a company operates in. Since Datastream merely provides up to ten business segments per company, we decided to group business sectors on a 3-digit SIC level. Table 1 illustrates that 66 of the 522 corporations concentrate their business activity in a single sector, while the remaining 456 companies are diversified across different sectors. Around half of the sample operates in four business segments or more. Considering the median  $D$  in the third column of Table 1, we observe a tendency for the diffusion coefficient to decrease with the number of business segments. To further quantify this impression, we have tested for differences between medians in the different groups. Comparing firms operating in one business segment with companies that are active in four (or more) sectors, a Mann-Whitney test rejects the null hypothesis that the average diffusion coefficient of non-diversified corporations is smaller or equal to that of diversified ones at the five percent level. The business segment count, however, lacks information on the relative importance of the different segments, that is on how much the respective revenues in these segments contribute to a corporation's overall sales. Therefore,

---

<sup>17</sup> Taking either the latest available observation of firm size or the maximum value per firm leads to very similar results.



Table 1: Median diffusion coefficient (and its square root) for firms that operate in a certain number of business segments (3-digit SIC level).

Business segment count	Number of firms	$D_{med}$	$\sqrt{D}_{med}$
1	66	0.001764	0.042
2	99	0.001681	0.041
3	97	0.001369	0.037
4	101	0.000841	0.029
5	70	0.000961	0.031
6	49	0.000784	0.028
7	29	0.001024	0.032
8	11	0.000625	0.025

*Note:* The persistence of abnormal profits seems to increase with the business segment count of a corporation.

Herfindahl suggests a diversification index that computes the sum of squared shares of each segment's contribution to total sales

$$H_i = \sum_{j=1}^n P_{ij}^2, \quad (18)$$

where  $P_{ij}$  is the percentage share of company  $i$ 's sales that is generated in business segment  $j$ . Notice that the measure decreases with increasing diversification. Alternatively, the entropy methodology can be applied to calculate a sales diversity index:<sup>18</sup>

$$E_i = - \sum_{j=1}^n P_{ij} \log P_{ij}. \quad (19)$$

Unlike the Herfindahl index, which weights the share of each business segment by itself, the entropy measure weighs each  $P_j$  by the logarithm of  $1/P_j$ , so that it is more sensitive to small sales shares than the Herfindahl index, and largely ignores small differences in large sectors. The entropy measure in equation (19) increases with increasing diversification.

The Spearman rank correlation coefficient for the relationship between  $\sqrt{D}_i$  and the Herfindahl-index is 0.17, indicating a moderate negative effect of diversification on the adjustment speed of the process.<sup>19</sup> Based on the Spearman rank test, we find that the

<sup>18</sup> For instance, Horowitz (1970) uses entropy as a measure of industry concentration, while Jacquemin and Berry (1979) use the concept to measure corporate diversification.

<sup>19</sup> Again, notice that more diversification leads to a reduction of the Herfindahl index, thus a positive correlation between  $D_i$  and the Herfindahl index implies that the speed of convergence toward the average profit rate decreases with increasing diversification.

null hypothesis of  $\sqrt{D_i}$  and the Herfindahl index being independent or negatively correlated can be rejected at the one percent level.

In case of the entropy measure, the correlation coefficient equals  $-0.15$  and the null hypothesis that  $\sqrt{D_i}$  and the entropy measure are not negatively correlated is also rejected at the 1 percent level. Furthermore, we have regressed the square root of the diffusion coefficient on these two diversification measures and find coefficients of  $0.0246 \pm 0.006$  for the Herfindahl-index and  $-0.0142 \pm 0.0038$  for the entropy measure. Both coefficients are statistically significant at the one percent level.

Overall, the results suggest that there is a moderate negative correlation between the diffusion coefficient and firm diversification, which is nevertheless quite robust with respect to several diversification measures. Abnormal profits would thus seem to be more persistent for more diversified corporations.

#### 1.4.5 Intensity of capital

The long relaxation times for insurance and utilities corporations indicate that capital intensity has an impact on the persistence of abnormal profits. According to the DuPont identity, the profit rate of a firm  $i$  can be decomposed into the product of its profit margin and asset turnover,

$$p_{i,t} = \frac{I_{i,t}}{S_{i,t}} \cdot \frac{S_{i,t}}{A_{i,t}}, \quad (20)$$

where (the inverse of) the latter measures capital intensity. The coefficient for the rank correlation between asset turnover and the noise level  $\sqrt{D_i}$  equals  $0.27$ , implying that abnormal profits are more persistent for capital intensive corporations (with a correspondingly low asset turnover). A one-sided Spearman rank test rejects the null hypothesis of nonpositive correlations at the one percent level.

Figure 9 provides an alternative illustration of the relationship between asset turnover and the noise level. We calculate the average asset turnover for each corporation during the period 1980-2011, group them into deciles, and calculate the median for each bin as well as the median of the associated noise levels  $\sqrt{D_i}$ . An ordinary least squares regression yields an intercept of  $0.0227 \pm 0.004$  and a slope parameter of  $0.0112 \pm 0.003$ , both of which are statistically significant at the one percent level. In summary, the abnormal profits of capital intensive corporations are more persistent.

## 1.5 DISCUSSION

The possibly most fundamental questions, pertaining to the origin of the particular values of the systemic rate of profit and its dispersion, still remain unanswered here. Yet the data suggest that statistical equilibrium provides a very reasonable first approximation to the profitability of surviving corporations.<sup>20</sup> Conditional on survival,

<sup>20</sup> We cannot emphasize enough that survival is key to our results, as already argued at length by Alfarano et al. (2012).

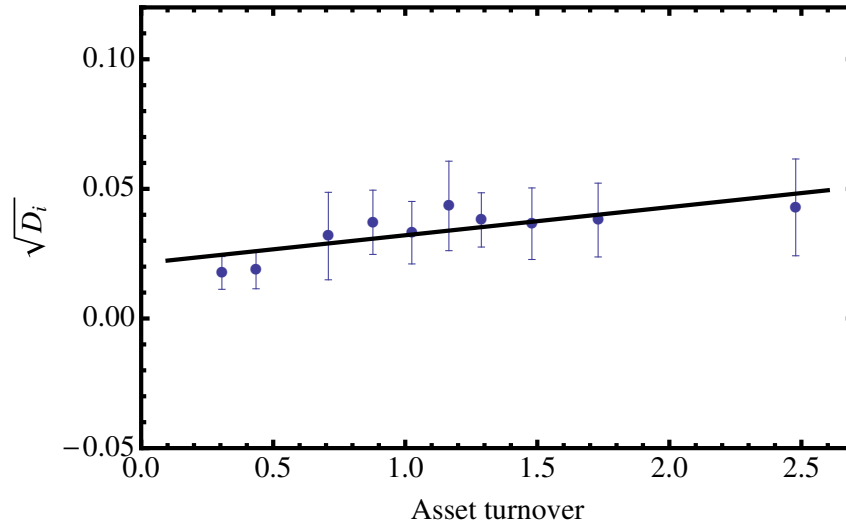


Figure 9: Estimates of the noise level ( $\sqrt{D}$ ) vs asset turnover, measured as the ratio of sales to total assets. Ordinary least squares regression of  $\sqrt{D_i}$  on asset turnover yields an intercept of  $0.0213 \pm 0.0040$  and a slope parameter of  $0.0109 \pm 0.0031$ . Both parameters are statistically significant at the one percent level; error bars represent one median absolute deviation.

US corporations generate an average rate of profit of about nine percent, along with a rather tranquil playground that disperses profit rates by less than six percent on average. Therefore it would appear that survival by itself warrants some sort of autopilot mode for corporations, in which they cannot do better but, perhaps surprisingly, also not worse than the system-wide average. Consequently, the idiosyncratic characteristics of corporations are independent of the systemic rate of profit and merely have an impact on how quickly abnormal profits are dissipated.

Knowledge of the relaxation time is of potential interest to outside investors because investing either in corporations which currently operate under a relatively large diffusion coefficient and are below the systemic rate of profit, or in those that currently operate above under a small diffusion coefficient, would both appear to be profitable strategies. These strategies, however, will pay off if and only if the respective corporations manage to survive, which is certainly unpredictable and therefore still leads to risky strategies.

As far as the much more abundant literature on growth rates is concerned, note that our sample builds on corporations from a broad set of industrial contexts, unlike previous studies that have mostly focused on particular sectors of the economy, such as manufacturing, services or the pharmaceutical industry. The results pertaining to the probability distribution, autocorrelation and scaling behavior of growth rates are consistent with the previous literature, in spite of the broader sectoral scope, again

indicating the importance of survival for the statistical regularities in both growth and profitability.

There are undeniably second-order effects that are not accounted for by the diffusion model. A substantial fraction of the deviations reported in sections 1.2 and 1.4 can be traced to a relatively small number of corporations in even fewer industries: about a dozen corporations with SIC codes 13 (oil and gas extraction), 36 (electronic equipment excluding computers), 38 (measuring instruments), and 67 (holding and other investment offices) account for half the deviations, which stem from the high volatility and leptokurtosis in the respective corporate time series. At the other end of the spectrum the largest corporations, like insurance carriers and utility companies, exhibit the least volatility and kurtosis in their profit rate series, and therefore account for most of the deviations in the relaxation time of abnormal profits. Recall that their relaxation times exceed the length of the observed series by almost an order of magnitude. So the idea that the most capitalized corporations are somehow privileged in a competitive environment (see, e.g., Baumol, 1967) lives on in a modified form, however not relating to the rate of profit itself or its volatility, but rather to the long persistence of abnormal profits in capital intensive industries.<sup>21</sup>

Pronounced deviations from the diffusion model might help to identify imperfections in the competitive environment, and potentially have antitrust implications. Since the diffusion model rests on the classical idea of a perpetual reallocation of capital in search of profit rate equalization, large empirical deviations in profit persistence should essentially be tied to frictions in the reallocation of capital.

Finally, and maybe most controversially, we would like to argue that concerns of a ‘survivorship bias’ are perhaps the wrong way of framing the empirical analysis. After all, Gabaix’s granular view of aggregate fluctuations in the US economy firmly suggests that the surviving corporations in our sample account for the major share of macroeconomic fluctuations, and are thus at the very least an interesting group of firms to study in its own right. Since everything in our analysis is conditional on (the uncertain and unpredictable) survival of corporations, one might instead wonder whether there is a systemic cost for the survival of a certain number of large corporations? Does capital need to be churned, do other corporations have to die, in order to observe the tranquil dissipation of excess profits for a certain (and ultimately interchangeable) set of surviving corporations? And if so, how much capital needs to be metabolized in the process?

---

<sup>21</sup> Notice that stakeholders in these industries, once their profitability actually happens to be *below* the systemic rate, might not consider it such a privilege after all.

## APPENDIX

## A.1 SAMPLE CHARACTERISTICS

Table 2: Sector definitions and number of firms in each sector.

Division	SIC	Sector	No. of firms	No. of long-lived firms
Agriculture, forestry, and fishing	01	Agricultural production - crops	20	1
	02	Agricultural production - livestock and animal specialities	3	0
	07	Agricultural services	4	0
	08	Forestry	3	0
	09	Fishing, hunting and trapping	2	0
Mining	10	Metal mining	139	4
	12	Coal mining	27	1
	13	Oil and gas extraction	336	15
	14	Mining and quarrying of nonmetallic minerals, except fuels	19	1
Construction	15	Building construction - general contractors and operative builders	37	5
	16	Heavy construction, except building construction - contractors	26	3
	17	Special trade contractors	21	1
Manufacturing	20	Food and kindred products	141	20
	21	Tobacco products	8	2
	22	Textile mill products	19	2
	23	Apparel, finished products from fabrics and similar materials	35	4
	24	Lumber and wood products, except furniture	25	5
	25	Furniture and fixtures	27	8
	26	Paper and allied products	36	8
	27	Printing, publishing and allied industries	71	15
	28	Chemicals and allied products	636	36
	29	Petroleum refining and related industries	31	10
	30	Rubber and miscellaneous plastic products	55	6

(continued)

Table 2 – Continued

Division	SIC	Sector	No. of firms	No. of long-lived firms
	31	Leather and leather products	19	3
	32	Stone, clay, glass and concrete products	30	6
	33	Primary metal industries	70	9
	34	Fabricated metal products	79	15
	35	Industrial and commercial machinery and computer equipment	297	36
	36	Electronic, electronic equipment and components	555	33
	37	Transportation equipment	147	24
	38	Measuring, analyzing and controlling instruments	422	33
	39	Miscellaneous manufacturing industries	59	4
	40	Railroad transportation	9	5
	41	Local, suburban transit and interurban transportation	3	0
Transportation, communications, electric, gas, and sanitary services	42	Motor freight transportation	30	3
	44	Water transportation	41	2
	45	Transportation by air	34	4
	46	Pipelines, except natural gas	10	0
	47	Transportation services	24	0
	48	Communications	227	8
	49	Electric, gas, and sanitary services	191	62
Wholesale trade	50	Wholesale trade-durable goods	145	10
	51	Wholesale trade-nondurable goods	100	11
	52	Building materials	8	1
	53	General merchandise stores	29	7
	54	Food stores	25	5
Retail trade	55	Automotive dealers and gasoline service stations	29	1
	56	Apparel and accessory stores	62	7
	57	Home furniture, furnishings, and equipment stores	28	3
	58	Eating and drinking places	85	7
	59	Miscellaneous retail	92	3
Finance, insurance and real estate	62	Security and commodity brokers, dealers, and exchanges	72	1
	63	Insurance carriers	141	20
	64	Insurance agents, brokers, and service	19	3

(continued)

Table 2 – Continued

Division	SIC	Sector	No. of firms	No. of long-lived firms
	65	Real estate	85	1
	67	Holding and other investment offices	334	11
	70	Hotels, rooming houses, camps, and other lodging places	42	1
	72	Personal services	15	4
	73	Business services	1018	19
	75	Automotive repair, services, and parking	12	1
Services	76	Miscellaneous repair services	3	1
	78	Motion pictures	54	0
	79	Amusement and recreation services	68	1
	80	Health services	118	5
	81	Legal services	2	0
	82	Educational services	49	1
	83	Social services	5	0
	84	Museums, art galleries, and botanical and zoological gardens	1	0
	86	Membership organizations	1	0
	87	Engineering, accounting, research, and management	187	4
	89	Services, not elsewhere classified	27	0
	92	Justice, public order, and safety	1	0
Public administration	95	Administration of environmental quality and housing programs	1	0
	97	National security and international affairs	8	0
	99	Nonclassifiable establishments	26	0
Total			6860	522

*Note:* Firms operating in more than one sector are classified according to the business segment that generated the most revenue. The fourth column refers to the whole dataset, while the fifth column represents long-lived firms.

## A.2 ESTIMATES OF THE SUBBOTIN PARAMETERS

Table 3 presents the estimated Subbotin parameters for profit and growth rates, distinguishing between different measures of firm size: besides total assets we also consider the sales, number of employees, and market value of corporations to check the robustness of our findings. Figure 2 in the main text already suggests that the Laplace fit is better for profit than for growth rates, which is confirmed by the estimates of the shape parameter  $\alpha$  since they are significantly smaller than unity for the growth of total assets, sales, and employment. The best Laplace fit is obtained for the growth rates of market value, yet an analysis of the year-by-year distributions (not shown here) reveals that the scale parameter exhibits pronounced fluctuations over time, suggesting that the Laplacian nature of the pooled market value growth rate distribution is an artefact of aggregation.<sup>22</sup>

Table 3: Maximum likelihood estimates of the Subbotin parameters  $\alpha$  and  $\sigma$ .

	shape parameter $\hat{\alpha}$	scale parameter $\hat{\sigma}$
profit rate	0.95 (0.01)	0.0570 (0.0005)
growth rate of total assets	0.76 (0.01)	0.0977 (0.0010)
growth rate of sales	0.74 (0.01)	0.0999 (0.0011)
growth rate of number of employees	0.62 (0.01)	0.0806 (0.0009)
growth rate of market value	1.01 (0.01)	0.2467 (0.0024)

*Note:* Standard errors are shown in parentheses.

## A.3 DISAGGREGATION

The analyses done so far considered firms from a diverse set of industries. A potential objection to this approach is that our results were affected by the aggregation of data, thus one referee suggested to check the validity of our findings in Section 1.2 for the firms in the manufacturing industries (SIC 20–39), where our sample contains 279 long-lived corporations. Hence this section replicates our analysis in Section 1.2 for these firms, shown in Figures 10–14. It turns out that the statistical properties observed across all sectors also hold within the manufacturing sector.

<sup>22</sup> This material is available upon request.



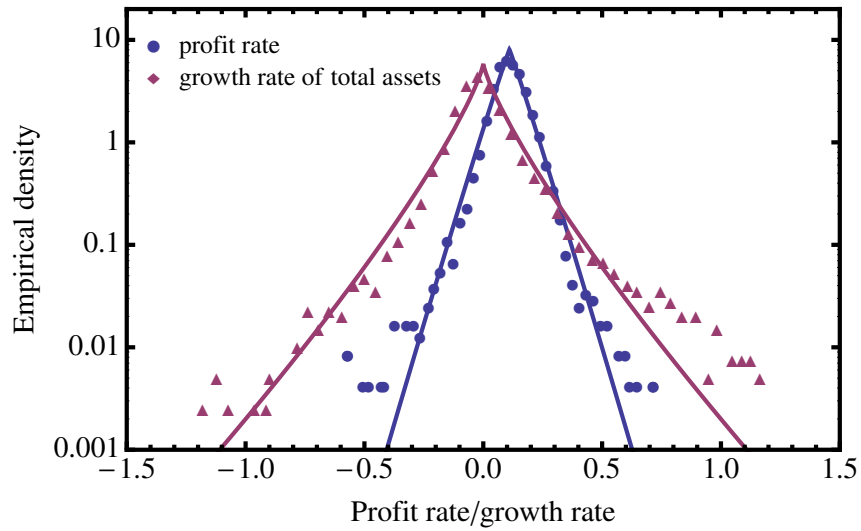


Figure 10: Pooled empirical densities of annual profit rates and growth rates for publicly traded manufacturing companies in the United States during the period 1980-2011. The mode of the profit rate distribution is  $\hat{m} = 11\%$ . Due to the normalization of the data, the growth rate distribution is centered around zero. The solid curves show the Subbotin fit obtained by maximum likelihood estimation of the shape and scale parameter, yielding the estimates  $\hat{\alpha} = 1.05$  (0.02),  $\hat{\sigma} = 0.0606$  (0.0008) for the profit rates and  $\hat{\alpha} = 0.80$  (0.01),  $\hat{\sigma} = 0.0990$  (0.0014) for the growth rates.

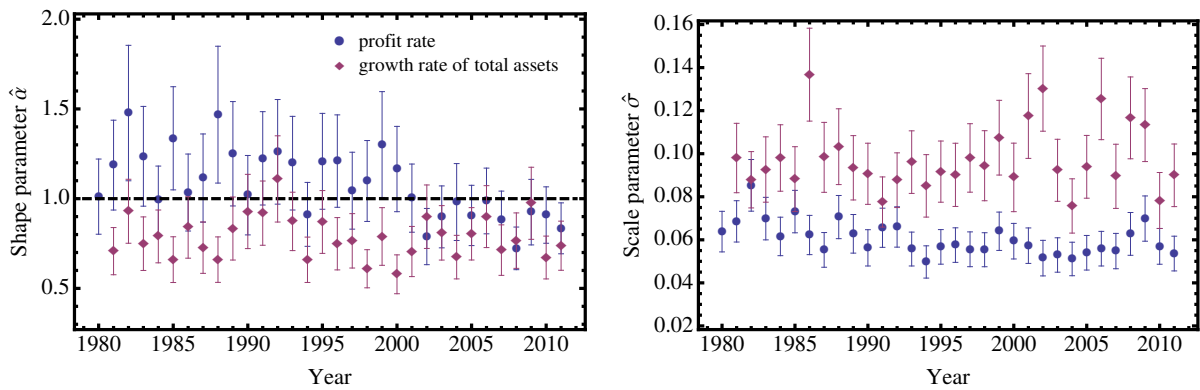
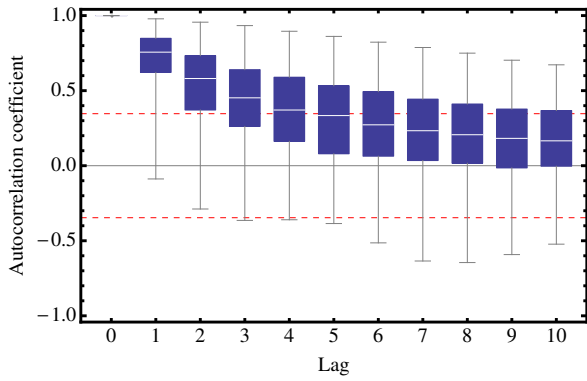
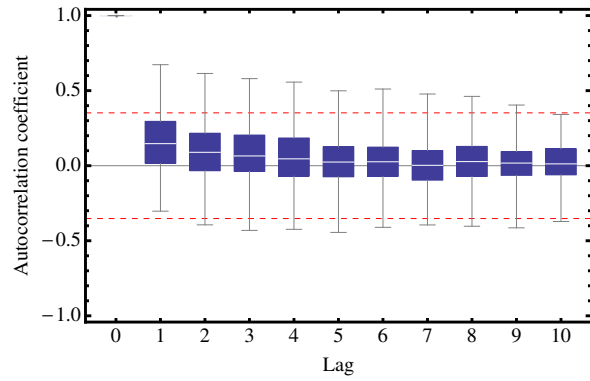


Figure 11: Year-by-year maximum likelihood estimates of the Subbotin shape (left panel) and scale (right panel) parameters for the firms in the manufacturing industries. Error bars show two standard errors.



(a) Profit rate



(b) Growth rate of total assets

Figure 12: Box-and-whisker plot for the estimated autocorrelation functions of profit rates and growth rates for firms in the manufacturing industries. The boxes include the 25 percent quantile, the median, and the 75 percent quantile. The red dashed lines show the 95 percent confidence interval under the null hypothesis of zero autocorrelations. The interval has been computed as  $\pm 1.96/\sqrt{T}$ , where  $T = 32$  is the length of the profit rate time series.

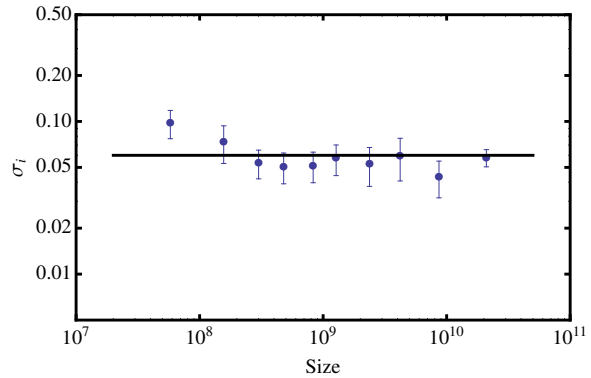
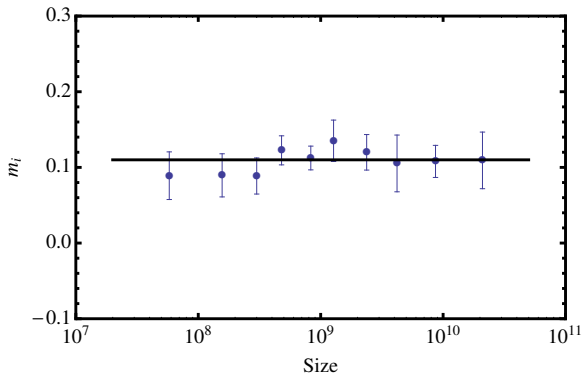


Figure 13: Location ( $m_i$ ) and dispersion ( $\sigma_i$ ) of profit rates as a function of firm size for firms in the manufacturing industries. The black lines represent the phenomenological  $\hat{m} = 11\%$  and  $\hat{\sigma} = 6.06\%$  of the empirical profit rate distribution. Error bars correspond to one median absolute deviation.

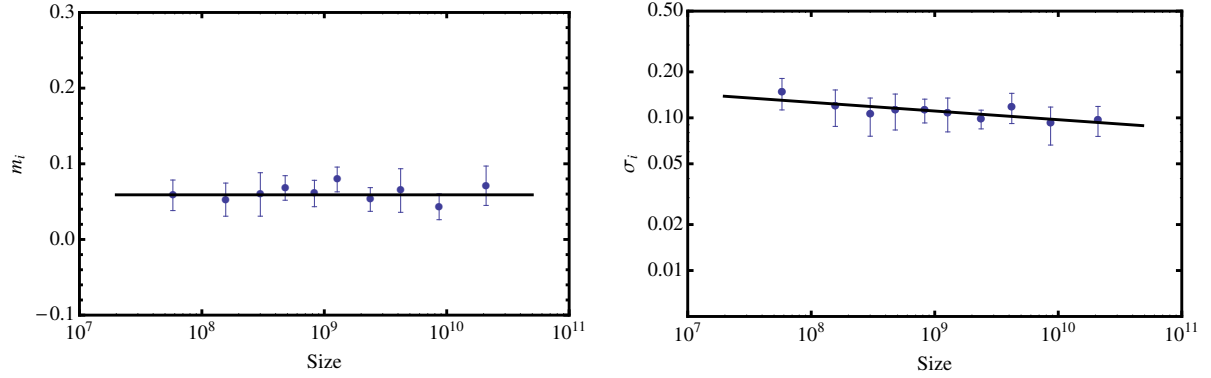


Figure 14: Location ( $m_i$ ) and dispersion ( $\sigma_i$ ) of growth rates as a function of firm size for firms in the manufacturing industries. A least squares power law fit for the relation between the dispersion of growth in total assets and size yields  $-0.06 \pm 0.02$ . Error bars correspond to one median absolute deviation.

#### A.4 SIMULATIONS

Simulations of the diffusion process (7) reproduce the stationary density and autocorrelation profile on a sufficiently long time scale, as illustrated in Figure 15. To demonstrate the relevance of the negative bias in the ACF for shorter time series, we have performed Monte Carlo simulations of the Laplace diffusion in equation (7) with different time series length  $T$ . We then calculated the autocovariance function by subtracting the sample mean from the observations. Figure 16 shows that subtracting the time series mean leads to a significant negative bias in the estimated autocorrelation function that becomes smaller once the length of the time series increases. In fact for shorter time series, as in our data ( $T \approx 30$ ), this particular shape of the autocorrelation function results from an exponentially decaying autocorrelation function and a bias which is linear in the time lag  $\tau$ . Subtracting the systemic profit rate  $m$  instead of each individual time series mean leads to a considerable reduction of the bias, but does not eliminate it completely. Thus we also computed the correlation time, based on the analytical solution of the autocorrelation function for our model, and report it as a more robust statistic in Section 1.4.1.

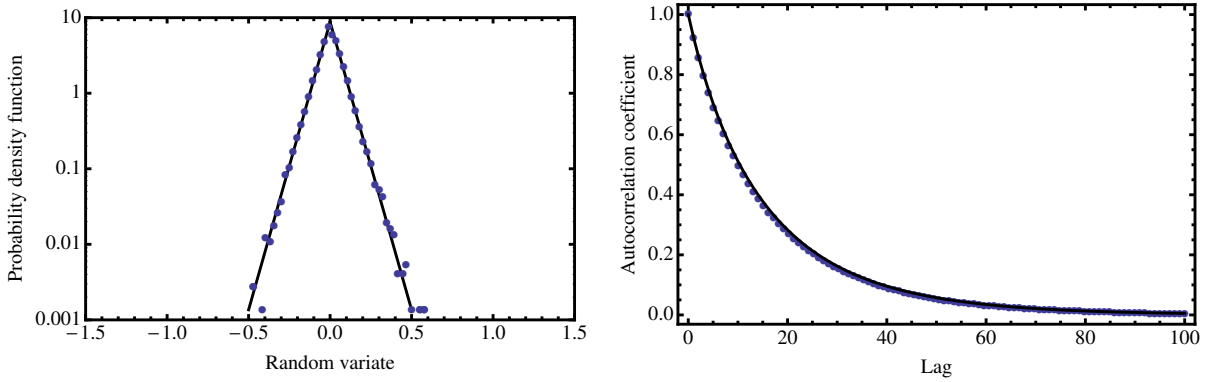


Figure 15: Simulation of the model by Alfarano et al. (2012). The left panel illustrates the probability density (blue dots) of 1000 simulated time series with common parameters  $\sigma = 0.057$ ,  $D = 0.001$ , and zero mean. The simulated data fit the probability density function of a Laplace distribution with corresponding location and scale parameters, which is represented by the black solid line. The right panel shows the estimated autocorrelation function of the diffusion process (blue dots) along with the theoretical autocorrelation function (black solid line). The estimated autocorrelation function has been averaged over 1000 realizations of the diffusion process with time series length  $T = 10000$ .

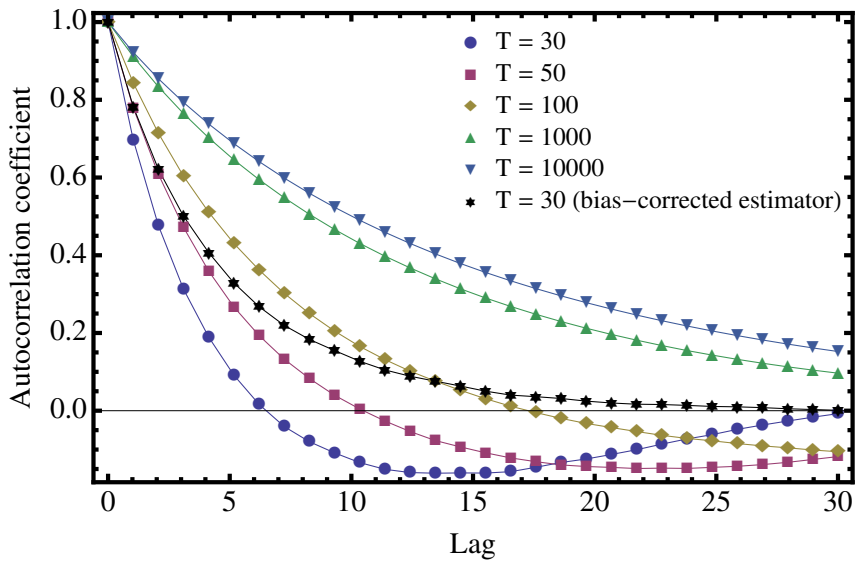


Figure 16: Autocorrelation function of the Laplace diffusion with parameters  $\sigma = 0.057$ ,  $m = 0$  and  $D = 0.001$  as a function of time series length  $T$ . Autocorrelation coefficients are calculated using the mean of each time series and are averaged over 1000 realizations of the process. For comparison, we also show the bias-corrected estimate that emerges when  $m$  replaces the time series mean in the autocovariance function (black stars).

## A.5 REGRESSION RESULTS

Table 4: Estimates for the slope coefficient in linear ordinary least squares regressions of the location parameter  $m$  on the logarithm of firm size.

Size measure	Slope including first bin	Slope excluding first bin
Profit rates		
Total assets	-0.0011 (0.0017)	-0.0030 (0.0019)
Sales	<b>0.0054</b> (0.0020)	0.0021 (0.0019)
No. of employees	<b>0.0072</b> (0.0025)	0.0045 (0.0029)
Market value	<b>0.0055</b> (0.0016)	0.0042 (0.0019)
Growth rates		
Total assets	-0.0002 (0.0016)	-0.0016 (0.0020)
Sales	-0.0010 (0.0012)	-0.0018 (0.0015)
No. of employees	0.0013 (0.0008)	0.0009 (0.0011)
Market value	<b>0.0107</b> (0.0029)	<b>0.0054</b> (0.0021)

*Note:* Standard errors are shown in parentheses. Bold values are statistically significant at the 5 percent level.

Table 5: Estimates for the slope coefficient in linear ordinary least squares regressions of the logarithm of the dispersion parameter  $\sigma$  on the logarithm of firm size.

Size measure	Slope including first bin	Slope excluding first bin
Profit rates		
Total assets	<b>-0.1015</b> (0.0303)	-0.0506 (0.0232)
Sales	<b>-0.0875</b> (0.0334)	-0.0210 (0.0184)
No. of employees	<b>-0.0831</b> (0.0301)	-0.0267 (0.0198)
Market value	-0.0602 (0.0361)	-0.0200 (0.0412)
Growth rates		
Total assets	<b>-0.0634</b> (0.0125)	<b>-0.0525</b> (0.0149)
Sales	<b>-0.1044</b> (0.0165)	<b>-0.0821</b> (0.0175)
No. of employees	<b>-0.0979</b> (0.0189)	<b>-0.0721</b> (0.0196)
Market value	<b>-0.0913</b> (0.0154)	<b>-0.0737</b> (0.0174)

*Note:* Standard errors are shown in parentheses. Bold values are statistically significant at the 5 percent level.



THE REAL VERSUS THE FINANCIAL ECONOMY:  
A GLOBAL TALE OF STABILITY VERSUS VOLATILITY

---

## ABSTRACT

The question how the real and the financial side of a capitalist economy relate to each other has been a frequently recurring topic in the history of economic thought. Our paper addresses this question from the viewpoint that capital ultimately seeks returns from its perpetual reallocation and essentially faces two choices: it can either be “entrepreneurially” allocated to real economic activity, or it can be “financially” invested in legal claims against such activity. Adopting such a perspective, we study here how real and financial returns relate to each other over the past fifteen years, both within and across countries, by considering more than 30,000 publicly traded firms in more than forty countries that stand for 70% of the global population and about 90% of world income. We compare the average rates of return to both types of investment and their respective volatilities. While average returns, perhaps somewhat surprisingly, turn out to be roughly equal across the two domains, the volatility of financial returns exceeds ‘real volatility’ by an order of magnitude. From a systemic point of view, these findings raise the question why capital would seek out financial investments in the first place.

## 2.1 INTRODUCTION

The question how the real and the financial side of a capitalist economy relate to each other has been a frequently recurring topic in the history of economic thought, and the call for papers of this special issue in *Economics* explicitly mentions its importance in light of the recent turmoils in the global economy. Our paper addresses this question from the viewpoint that capital ultimately seeks returns from its perpetual reallocation and essentially faces two choices: it can either be “entrepreneurially” allocated to real economic activity (that is the production of goods and services), or it can be “financially” invested in legal claims against such activity. Adopting such a perspective, we study here how real and financial returns relate to each other over the past fifteen years, both within and across countries, by considering more than 30,000 publicly traded firms in more than forty countries that stand for 70% of the global population and about 90% of world income. We compare the average rates of return to both types of investment and their respective volatilities. While average returns turn out to be roughly equal across the two domains, the volatility of financial returns exceeds ‘real volatility’ by an order of magnitude. We also find that real returns are positively autocorrelated and exhibit remarkable stability over time, while financial returns have no memory and are characterized by pronounced fluctuations that are hard to reconcile with fluctuations in the real returns to economic activity. From a systemic point of view, these findings raise the question why capital would seek out financial investments in the first place.

Our perspective owes its intellectual debt to at least three influences. Firstly, we take the position that the destinies of the largest firms in an economy are of crucial quantitative importance for aggregate outcomes, a viewpoint that Gabaix (2011) calls the “granular origins of aggregate fluctuations”. Secondly, we focus on the profit



rate as a measure of the returns to real economic activity, an idea that has pervaded classical economic thinking since the times of Adam Smith (see, e.g., Foley, 2006). Last but not least, the latent notion of some form of excess volatility in financial returns dates back to the seminal contributions of Shiller (1981) and LeRoy and Porter (1981).

The pre-analytical vision of our study rests on the observation that positive long-run deviations of financial returns from the real rate of profit amount to a Ponzi scheme and are therefore unsustainable, simply because in the long run companies cannot afford to pay more to financial stakeholders than they earn from their real activities. On the other hand, negative long-run deviations coupled with arbitrage considerations would render financial investments relatively unattractive. Two important strands of literature have helped to shape this way of thinking, the first being the fundamental principle of valuation put forth by Miller and Modigliani (1961), the second being the efficient markets hypothesis of Fama (1970, 1991). According to the first, deviations from a firm's fundamental share price, perceived as the discounted value of future income streams, should be eliminated through trading on perfect capital markets. More importantly, Modigliani and Miller argue that it is ultimately the left-hand side of a company's balance sheet that matters, i.e. its real activities, and not how the right-hand side of the balance sheet is composed, i.e. how exactly and in which proportions the different financial claims are stacked against these real activities. According to the second, price movements are directly tied to the arrival of new information about activities in the real economy, implying that financial returns should be coupled to returns in the real economy. We find it unfortunate, however, that both these strands as well as the excess volatility literature have apparently singled out corporate dividends to be the most important determinant of a company's income stream. The reason why we find this choice less than ideal is that (i) dividend policy varies widely across companies and that (ii) the pertinent literature still, after more than half a century, has not determined a robust effect of dividends on share prices (see, e.g., Al-Malkawi et al., 2010, for a recent review of the field).

In order to compare real and financial returns, we propose to consider a firm's (possibly negative) ratio of its operating income to its total assets as a meaningful measure of the profit rate, in line with the orthodox accounting and business economics literature where it is also known as the return on assets, or ROA. Our measure of financial returns will be the (possibly negative) growth rate of the corresponding firm's financial market value. Choosing these two proxies for real and financial returns of course still represents an imperfect and stylized approach, mainly because the ROA is influenced by empirical accounting issues (see, e.g., Burgstahler and Dichev, 1997) and because the growth rate of market value does not implicitly account for the peculiarities of a publicly traded company's dividend policy. Yet we would like to believe that both quantities are useful first-order approximations to the real and financial rates of return.

Interestingly, the profit rate has not been at the forefront of economic inquiry for many decades, which is probably due to the fact that orthodox theories have not been able to provide a coherent explanation for the profit rate (see, e.g., Naples and Aslan-

beigui, 1996). Profit rates have nevertheless been studied in the context of the so-called persistence of profits literature that starts with Mueller (1977) (see, e.g., Cuaresma and Gschwandtner, 2006, for a more recent take on the subject), and more recently also in the context of a statistical equilibrium framework Alfarano and Milaković (2008) and Alfarano et al. (2012, 2013) that will guide our present investigation.<sup>23</sup>

## 2.2 DATA DESCRIPTION AND SAMPLE SELECTION

The data used in this study come from Thomson Reuters Datastream and consist of annual observations on operating income, total assets and market value for 32,201 publicly traded domestic companies from 43 different countries listed in Table 6. The countries in our sample stand for approximately 70 percent of world population and represent the largest economies in terms of world income, accounting for more than 87 percent of global gross domestic product in 2011 according to the IMF's World Economic Outlook database. The dataset contains firms which have been present in the market for at least one year between 1997 and 2011. It is filtered according to two criteria: first, we exclude banks (entities with SIC codes 60 and 61 on a two-digit classification level) from the analysis because it is well known that their balance sheets and profit rates differ from those of non-banks by at least one order of magnitude. Second, to check to what extent our results are affected by the entry and exit of firms, we create two different samples. The first one considers entities that report data on all three variables (operating income, total assets, and market value) in at least one period, hence it includes firms with life spans varying between one and fifteen years. The second sample focusses on long-lived or "surviving" firms that we define as companies reporting data in the first and the final period of the time window. The time period has been chosen in such a way to maximize the number of observations across countries.<sup>24</sup>

Table 6 provides information about the number of firms in both the entire sample and the subset of firms that fulfill the longevity criterion, comprising 7,488 surviving firms. Datastream does not contain enough companies in Egypt and Russia that can be classified as long-lived companies according to our criterion, thus we only present results for firms with shorter life spans for these two countries. Averaging across all countries, around one quarter of the firms can be classified as long-lived. Notice, however, that these surviving firms on average account for approximately 60 percent of a country's total market capitalization according to World Bank data and, therefore, must be regarded as an important driver of economic activity. This argument is also supported by Gabaix (2011) who finds that about one third of variations in US GDP growth can be attributed to the idiosyncratic destinies of the largest one hundred US firms.

---

<sup>23</sup> Foley (1994) and Garibaldi and Scalas (2010) provide useful background material for readers who might not be entirely familiar with the concept of statistical equilibrium. To the best of our knowledge, Farjoun and Machover (1983) provide the first probabilistic perspective on the rate of profit.

<sup>24</sup> Datastream provides the most extensive coverage for US firms, going back to 1980. For most countries in our sample, however, coverage only begins in 1997.

Table 6: Countries under consideration.

Country	Number of firms	Number of long-lived firms	Survival rate (in %)
Argentina	74	26	35.1
Australia	1700	128	7.5
Austria	81	35	43.2
Belgium	138	44	31.9
Brazil	442	78	17.6
Canada	187	104	55.6
Chile	191	62	32.5
China	2053	153	7.5
Czech Republic	15	4	26.7
Denmark	155	76	49.0
Egypt	107	0	0.0
Finland	128	57	44.5
France	727	228	31.4
Germany	952	233	24.5
Greece	268	81	30.2
Hong Kong	1199	298	24.9
India	2249	222	9.9
Indonesia	358	120	33.5
Ireland	35	16	45.7
Israel	420	21	5.0
Italy	260	83	31.9
Japan	3378	1589	47.0
Malaysia	851	227	26.7
Mexico	128	50	39.1
Netherlands	107	67	62.6
New Zealand	135	32	23.7
Norway	178	39	21.9
Pakistan	162	48	29.6
Poland	381	13	3.4
Portugal	55	20	36.4
Russia	361	1	0.3
Singapore	694	123	17.7
South Africa	321	69	21.5
South Korea	1614	235	14.6
Spain	130	53	40.8
Sweden	421	91	21.6
Switzerland	217	111	51.2
Taiwan	1459	168	11.5
Thailand	768	220	28.6
Turkey	309	65	21.0
United Kingdom	1353	424	31.3
United States	7411	1770	23.9
Venezuela	29	4	13.8
Total	32201	7488	23.3

*Note:* The numbers in the second and third column refer to non-bank companies. Firms for which Datastream does not provide data on all three variables have been removed. The survival rate in the fourth column is computed as the ratio of the number of long-lived firms to the number of all firms in that country.

Based on these data we compute two quantities for each firm: the profit rate or return on assets, and the growth rate of market value. The profit rate of company  $i$  in year  $t$  is computed as the ratio of operating income ( $I$ ) to total assets ( $A$ )

$$p_i(t) = \frac{I_i(t)}{A_i(t)}, \quad (21)$$

while the growth rates are computed as logarithmic time differences in market value (MV)

$$g_i(t) = \log(MV_i(t+1)) - \log(MV_i(t)), \quad (22)$$

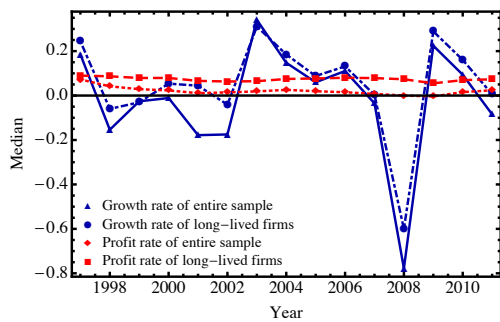
which should approximate the annual percentage change in market value. Notice that the comparison of both quantities is dimensionally sound in the sense that both quantities measure the annualized return per invested capital.

### 2.3 EMPIRICAL RESULTS

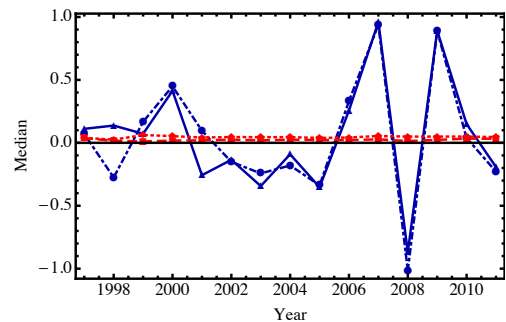
Since our goal is to obtain a global perspective of real and financial rates of return, we will focus on the time series properties of cross-sectionally averaged real and financial rates of return. Hence, our focus shifts from the return of a single stock or company to the central locations of the profit rate and growth in firm market value distributions in a given country. In the following analysis, we employ the median as location parameter since it is a more robust estimator against outliers.

Figure 17 illustrates the time evolution of the median profit rate as well as the median financial return for the six largest economies in the world: the US, China, Japan, Germany, France, and Brazil. The diagrams for the remaining countries are provided in Figures 21–26 in the appendix. For all countries we observe pronounced differences between the real and the financial side of the economy regarding the intensity of market reactions. Although there are also moderate up- and downturns in the median profit rate (notice for instance the decline in firm profitability in the course of the recent financial and banking crisis in almost every considered market), the rate of profit exhibits a remarkable stability over time that is at odds with the high volatility in financial returns. This non-trivial stability of the profit rate has already been pointed out by Alfarano et al. (2012), who study the distributional details of profit rates for the US for a time span that dates back to 1980 and is thus about twice as long as the present one. One of their findings is that the average rate of profit (measured for instance by the mode or median of the profit rate distribution) exhibits the same stability that we find here. Hence we would like to believe that our present results are not an artefact of the chosen time period.

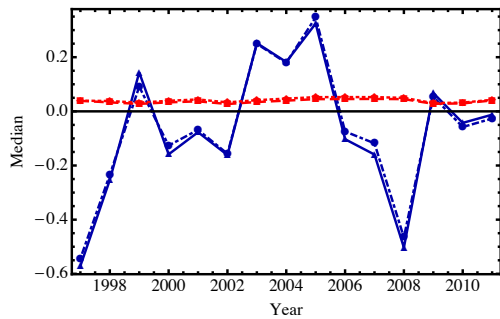
From an economic point of view, Alfarano et al. (2012), argue that the considerable stability of the profit rate should stem from the notion of classical competition that gives rise to a negative feedback mechanism, whereby capital seeks out sectors or industries where the profit rate is higher than the economy-wide average, typically attracting labor, raising output, and reducing prices and profit rates in the process.



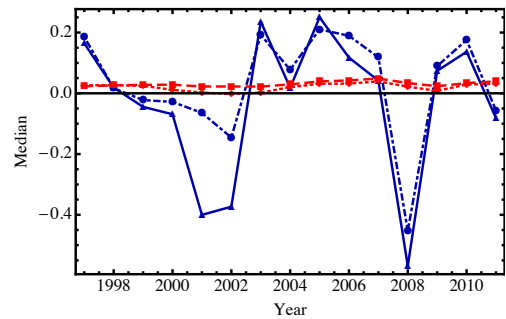
(a) United States



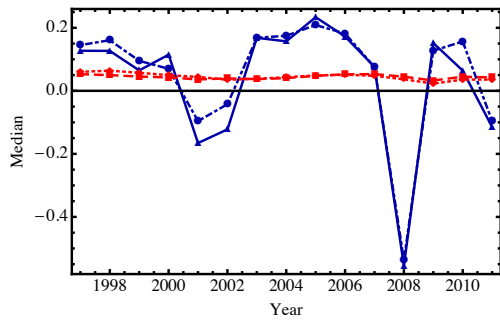
(b) China



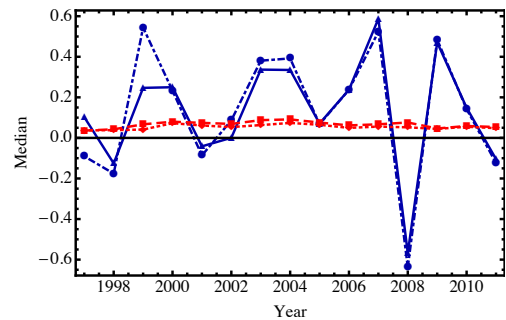
(c) Japan



(d) Germany



(e) France



(f) Brazil

Figure 17: Time evolution of the median profit rate and the median growth rate of market value for the United States, China, Japan, Germany, France, and Brazil. Results are shown for the entire sample and the long-lived firms. For visual clarity, linear interpolations between annual data points have been added.

This in turn provides an incentive for capital to leave the sector again, leading to higher prices and profit rates for firms that remain in the industry.

On the other hand, the growth rates of market value appear to fluctuate around the rate of profit, but their volatility is far too large to be explained by changes in the return to real economic activity. Hence, instead of the negative feedback mechanism that characterizes the real sector, financial markets seem to be subject to some sort of positive feedback mechanism and strong cross-correlations that drive stock prices into the same direction for extended periods of time. In fact, the recent literature on heterogeneous agent-based financial market models agrees in all its different flavors (see, e.g., Lux et al., 2007, for a comprehensive and fairly recent review) that positive feedbacks, typically in the form of self-reinforcing social interactions, are crucial for the reproduction of the observed stylized facts regarding the statistical properties of financial returns.

Visual inspection of the median time series also suggests that returns in the real economy are more persistent than financial returns. To quantify this impression, we have calculated the first-lag autocorrelation coefficient for the median profit rate and growth in market value series, using the estimator

$$\gamma = \frac{1}{T} \sum_{t=1}^{T-1} (X_t - \bar{x}_T)(X_{t+1} - \bar{x}_T), \quad (23)$$

where

$$\bar{x}_T = \frac{1}{T} \sum_{t=1}^T X_t \quad (24)$$

is the mean of  $T = 15$  observations from the time series. The results presented in Figure 18 support the view that annual returns earned in the real economy are positively autocorrelated, while there are no statistically significant autocorrelations in growth rates of market value, in line with the (weak-form) efficient market hypothesis.<sup>25</sup> This finding is very well established in the literature (see, e.g., Cont, 2001, for a review of the empirically established statistical properties of financial returns).<sup>26</sup> In light of the behavior of financial returns, the fact that autocorrelations in profit rates persist for one year (and most probably even longer) appears striking to us. We very much suspect that this finding traces back to real frictions and inertia introduced by, for instance, barriers of entry, the need to create and maintain corporate infrastructure, the administrative burden of founding a company, or the efforts and costs involved in hiring and releasing employees, which are all absent from financial capital investment.

<sup>25</sup> Note that the short length of the time series introduces a negative bias in the estimated autocorrelation coefficient (see, e.g., Fuller, 1996). Therefore, the autocorrelation in profit rates is even stronger than Figure 18 suggests. This bias probably explains the negative but statistically insignificant estimates for the autocorrelation coefficients of market value growth rates.

<sup>26</sup> If at all, significant autocorrelations in financial returns can merely be found on much smaller intraday time scales (so-called high frequency data) for which microstructure effects come into play.

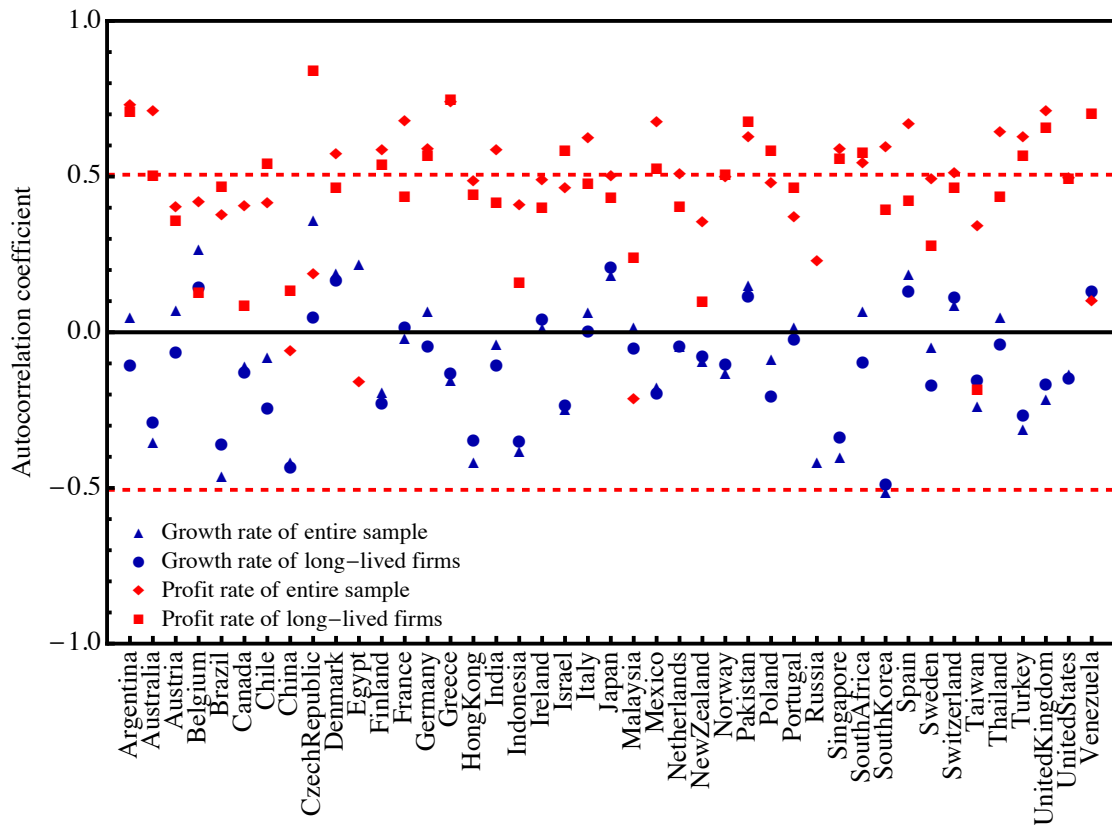


Figure 18: First-lag autocorrelation coefficients of both the median profit rate and growth in market value series for all countries in our sample. The red dashed lines show the 95% confidence interval under the null hypothesis of zero autocorrelations. The interval has been computed as  $\pm 1.96/\sqrt{T}$ , where  $T = 15$  is the length of our time series.

Moreover, our analysis indicates that there are differences in average firm profitability across countries. If we compute the median of the median profit rate series, the results for the long-lived companies vary between 1 percent in case of Portugal and approximately 10 percent for Pakistan. When the entire sample is considered, on one end Pakistan still has the highest average profit rate (approximately 9.7 percent), but now a high incidence of negative reported earnings in Australia at the other end of the spectrum results in an average return on assets of -6.1 percent (see Figure 27 in the appendix). This raises the question whether markets with a high profit rate are also more attractive to financial investors. Figure 19 presents a scatter plot showing combinations of the profit rate and the growth rate of market value (both averaged over the time series and across firms) for all countries in our long-lived sample. A weighted linear least squares regression of the financial return on the rate of profit that takes into account differences in market size or economic importance across countries is

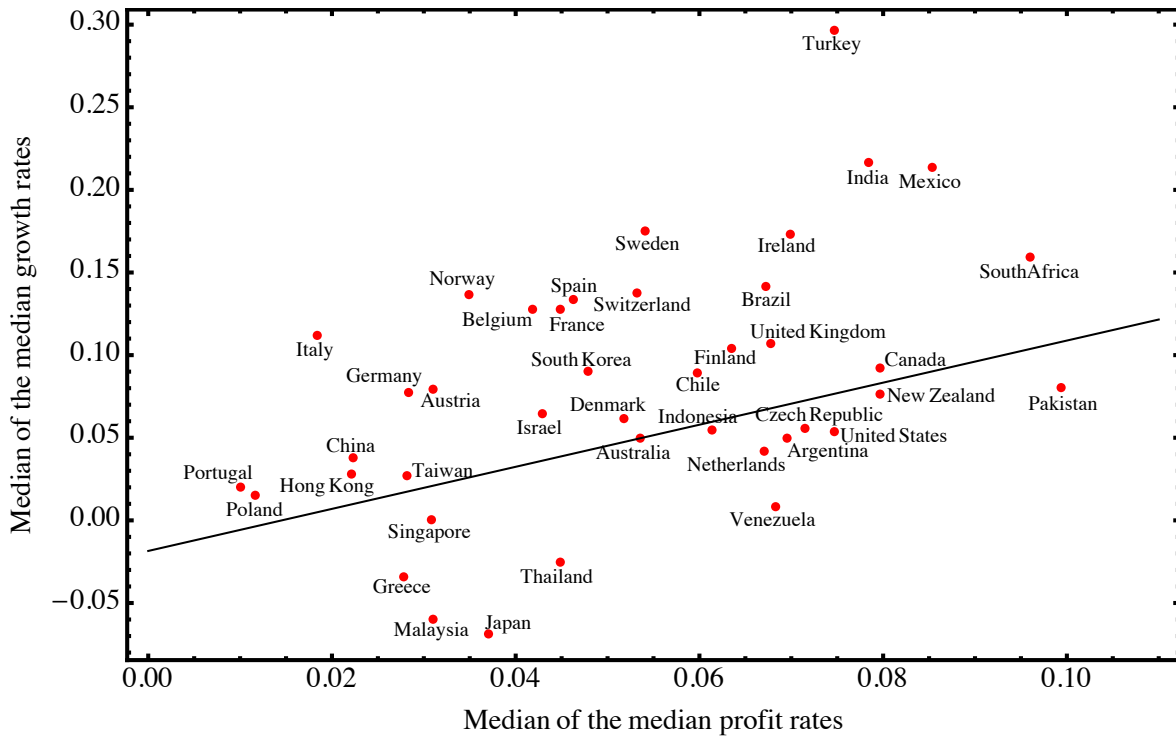


Figure 19: Scatter plot showing combinations of the average profit rate and the average growth rate of market value for long-lived firms in each country. The data points have been computed as the median of the median time series. Weighted least squares regression of the average financial return on the average profit rate yields an intercept of  $-0.02 \pm 0.03$  with a p-value of 0.57 and a slope parameter of  $1.27 \pm 0.52$  with a p-value of 0.02. Thus we cannot reject the hypothesis that average real and financial returns are the same (that is, a slope parameter of unity) at the usual confidence levels. The weights have been calculated by starting from the sum of market values of all (long-lived) firms in a country relative to global market capitalization for a given year, and have then been averaged over the period 1997-2011.



also shown.<sup>27</sup> We chose the weighting function to be the sum of market values of all (long-lived) firms in a country as a percentage of the market capitalization of the entire sample. To avoid distortions arising from booms and busts in single years, we calculate the weight for every year between 1997 and 2011 and then take the mean of these 15 values for each country. The weighted regression yields a slope coefficient of  $1.27 \pm 0.52$  with a p-value of 0.02. Since the estimate for the slope coefficient cannot be statistically distinguished from unity at the usual confidence levels, we take this to imply that investments in the real and the financial sector yield the same return *on average*. Thus, we may conclude that at least on the aggregate level financial returns are tied to the rate of profit, supporting the hypothesis that the profit rate is an important benchmark for financial returns *on average*. Our results also carry over to the entire sample of firms, in which case we find an intercept of  $-0.02 \pm 0.02$  with a p-value of 0.38 and a slope of  $1.42 \pm 0.54$  with a p-value of 0.01 (see Figure 27 in the appendix). Next we conduct a similar exercise for the volatility of the two quantities, measured as the median absolute deviation of the median time series. We chose this particular dispersion measure because it is more robust against outliers than the standard deviation.

Figure 20 presents the results for the long-lived companies. The scatter plot for the entire sample is provided in the appendix (Figure 28). We observe in both samples that the volatility of financial returns is about one order of magnitude higher than the volatility of profit rates, confirming the visual impression from the time series plots that financial returns are “excessively volatile” compared to profit rates. To check whether the two volatilities are related, we have regressed the median absolute deviation of growth in market value on the volatility of profit rates, again weighting countries with their percentage share of total market capitalization. However, in contrast to our results for the median, we do not find any clear relationship between the two variables. A weighted least squares regression for the set of long-lived firms yields an intercept of  $0.08 \pm 0.03$  with a p-value of  $4 \times 10^{-3}$ , but the slope parameter of  $6.36 \pm 4.01$  only has a p-value of 0.12. Regressing the volatility of profit rates on the volatility of financial returns and weighting the data with total assets instead of market values, we obtain a constant of  $0.01 \pm 0.001$  with a p-value of  $4.8 \times 10^{-7}$  and a slope of  $0.01 \pm 0.01$  with a p-value of 0.12. Similar results are found for the entire sample, and we are happy to provide them upon request. Therefore, although there is a relation between the two rates of return in terms of the median, the fluctuations in financial returns seem to be disconnected from fluctuations in the return of real economic activity.

As we argued before, one popular explanation for the excess volatility in financial returns and the endogenous dynamics of the financial sector that seem to be “disconnected” from fundamental factors are speculative activities of traders and the presence of herding behavior in financial markets, but not in the real sector. Since in both volatility regressions the slope is not significantly different from zero, the estimator for the intercept can be interpreted as a (weighted) sample average of volatilities

---

<sup>27</sup> An (unweighted) ordinary least squares regression leads to similar results and supports our central findings.

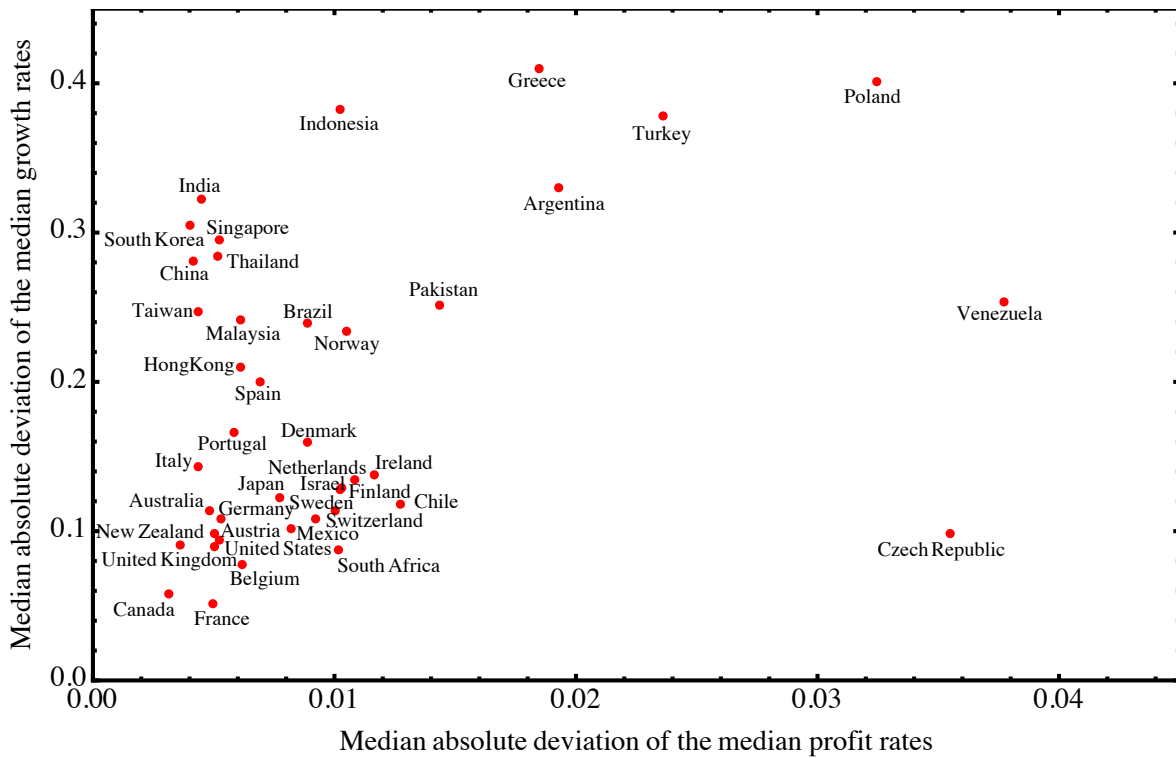


Figure 20: Scatter plot showing combinations of the profit rate volatility and the volatility of growth rates of market value for long-lived firms in each country. The data points have been computed as the median absolute deviation of the median time series. It is noteworthy that the volatilities differ by one order of magnitude, and that we cannot reject the hypothesis that the slope coefficient in a linear regression is equal to zero at the usual significance levels.

across countries. Thus the estimates confirm the impression that volatility in financial returns exceeds volatility in profit rates by about one order of magnitude.

## 2.4 DISCUSSION AND CONCLUDING REMARKS

While the length of the available time window in Datastream is certainly not ideal, the period 1997 to 2011 nevertheless strikes us as instructive for two reasons. It contains a period of considerable “financialization” or “securitization” of the global economy that starts in the 1990s, but it also contains a period of substantial financial distress through the global economic and banking crises that began in 2007. One might wonder whether the average equality of profit rates and financial returns would also hold without this realignment period? Reproducing the median regressions (Figures 19 and 27) for the period 1997-2006, we find a slope coefficient of  $1.59 \pm 0.53$  for the long-lived corporations and  $1.44 \pm 0.66$  for the entire sample (both parameters are statistically significant at the 5% level), thus our results for the years prior to the crisis are consistent with those reported in section 2.3 for the entire sample period, and we reject the hypothesis that the average equality of returns is merely due to the presence of the most recent crisis in our sample. Overall, our findings are compatible with some form of “investor rationality” since returns appear to be the same *on average*, so that investors eventually realize that irrational exuberances or panics cannot last forever.

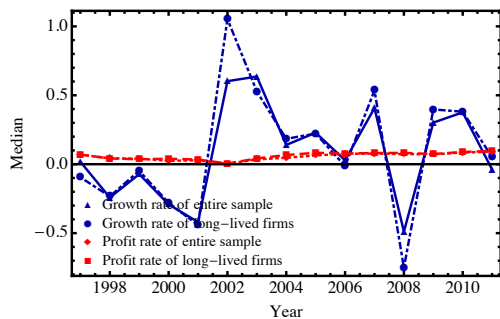
On the other hand, the source of deviations in volatility is most likely due to negative versus positive feedback mechanisms in the operation and allocation of real and financial capital. From our point of view, this leads to the question why capital would seek out financial market allocations in the first place. At this point, we find the observation by Shackle (1967) instructive, who claims that the foremost purpose of financial investment (or ‘money’ as he called it in the 1960s) “is the refuge from specialized commitment, the postponer of the need to take far-reaching decisions” because it provides much higher liquidity compared to the commitment of capital to real activity.

This view has intuitive appeal, but then one ultimately has to confront the question whether the possibility of postponing specialized commitments comes at a macroeconomic cost. Interest in this question dates back to the work of Kaldor (1956) and Pasinetti (1962), who have put forth what is often termed the *Cambridge growth equation*, a theory that in a more contemporary language boils down to statements about the relationship between the profit rate and the financial rate of return (see, e.g., Ciccarone, 2004, for a recent take on the subject), with far reaching implications for the functional distribution of income and macroeconomic stability at large. Orthodox interest in this subject has seemingly vanished altogether, which is probably due to the critique of the Cambridge growth equation by Samuelson and Modigliani (1966). In retrospect this strikes us as a rather unfortunate development, particularly since Kaldor, Pasinetti and Robinson have argued in their replies to Samuelson and Modigliani (that were published in the same issue) that the “Anti-Pasinetti” critique would require labor’s propensity to save to become so high as to allow the accumu-

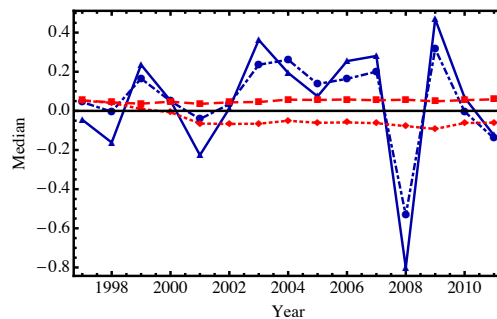
lation of capital through labor at a rate that is greater than the speed at which capitalists accumulate capital. But then the total capital of the economy would eventually be entirely owned by workers, while the capitalists would become extinct. Casual observation of economic history suggests that such an outcome does not appear to be very likely.

Irrespective of these long-standing theoretical debates, we would like to conclude by pointing out once more that from a macro-perspective we find it most surprising that the profit rate appears as such an enormously stable and positively autocorrelated variable in each country, making it a very worthwhile candidate for further study in our opinion, despite the apparent orthodox disinterest in the subject.

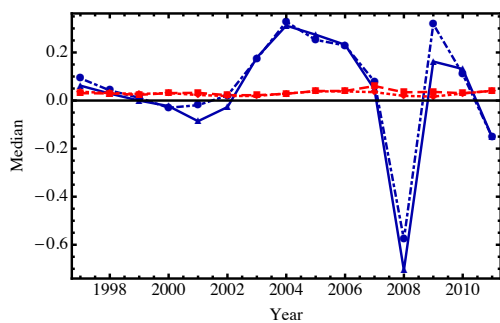
## APPENDIX



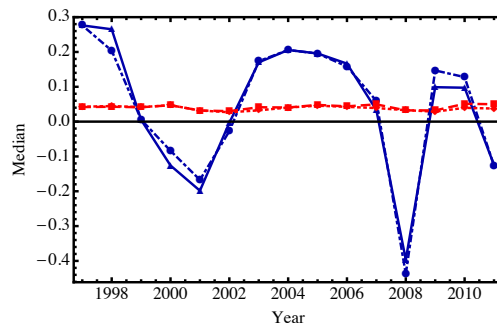
(a) Argentina



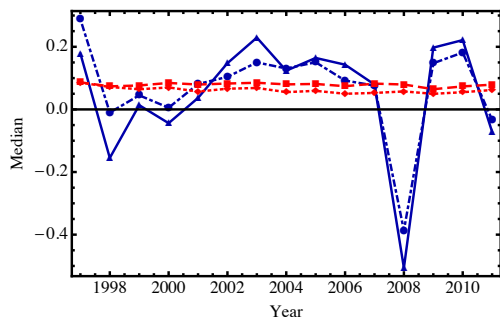
(b) Australia



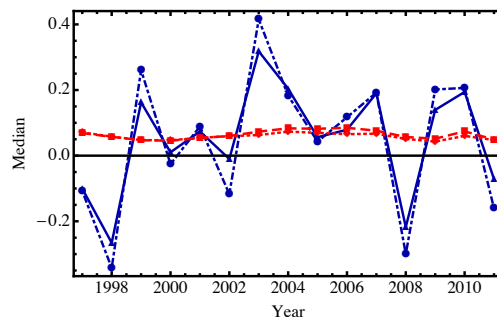
(c) Austria



(d) Belgium

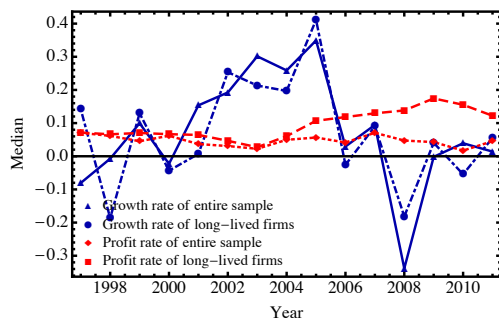


(e) Canada

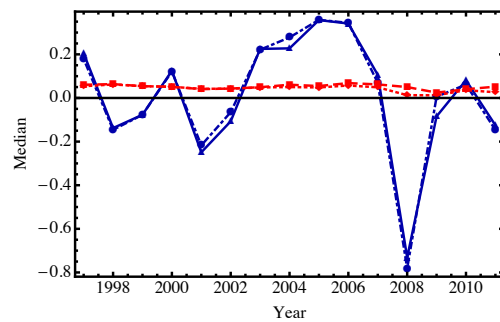


(f) Chile

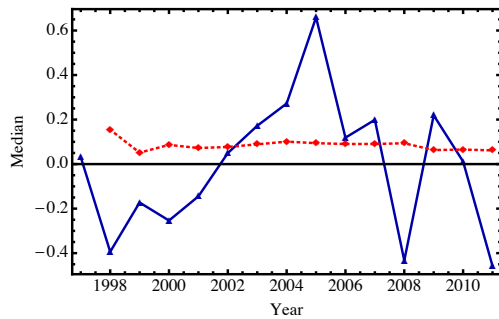
Figure 21: Time evolution of the median profit rate and the median growth rate of market value for Argentina, Australia, Austria, Belgium, Canada, and Chile. Results are shown for the entire sample and the long-lived firms. For visual clarity, linear interpolations between annual data points have been added.



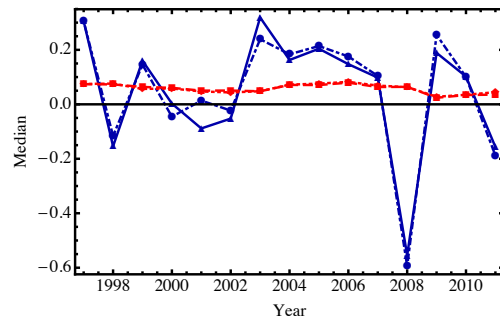
(a) Czech Republic



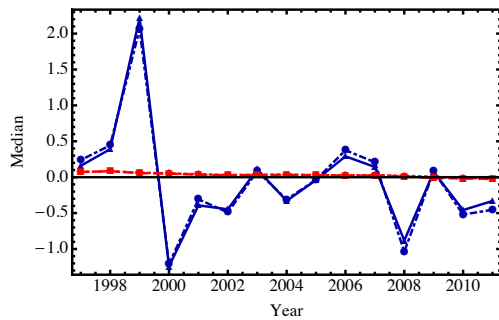
(b) Denmark



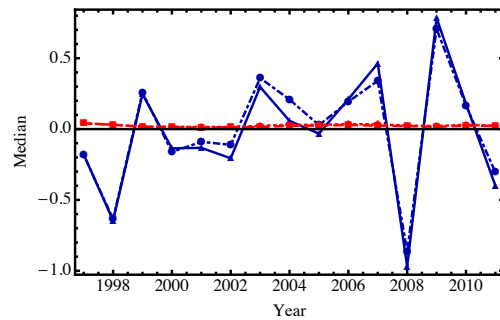
(c) Egypt



(d) Finland

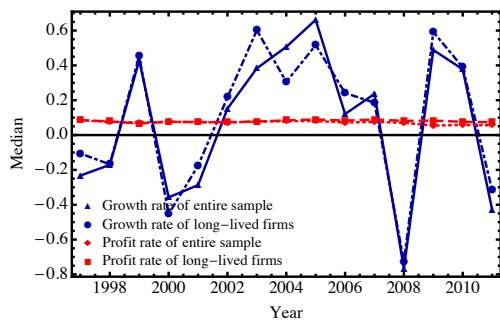


(e) Greece

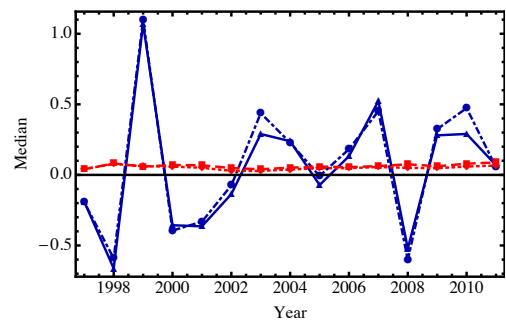


(f) Hong Kong

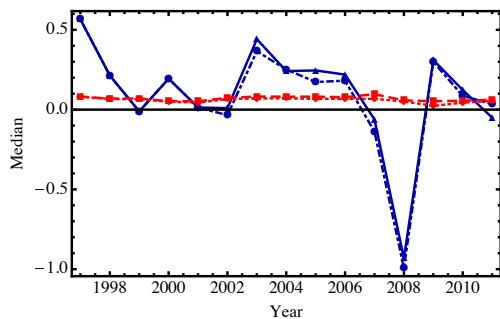
Figure 22: Time evolution of the median profit rate and the median growth rate of market value for Czech Republic, Denmark, Egypt, Finland, Greece, and Hong Kong. Results are shown for the entire sample and the long-lived firms. For visual clarity, linear interpolations between annual data points have been added.



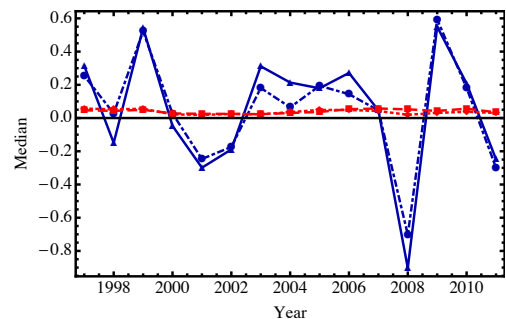
(a) India



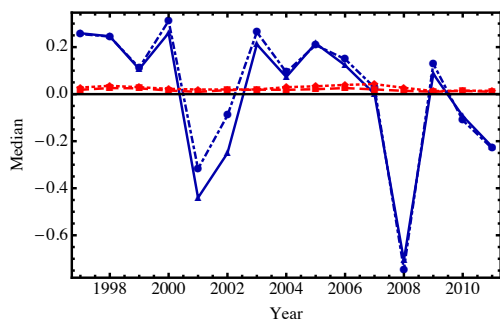
(b) Indonesia



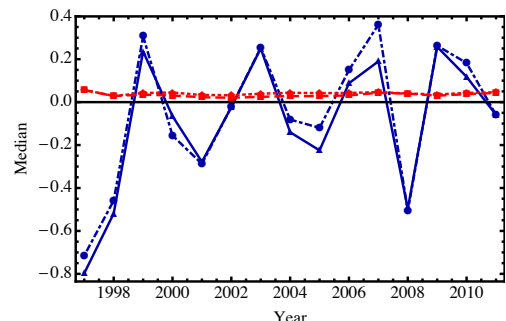
(c) Ireland



(d) Israel

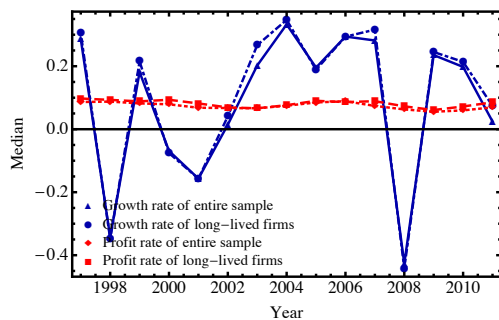


(e) Italy

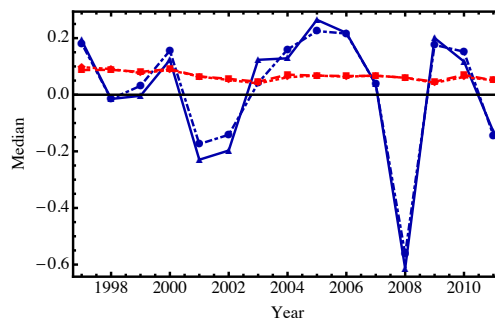


(f) Malaysia

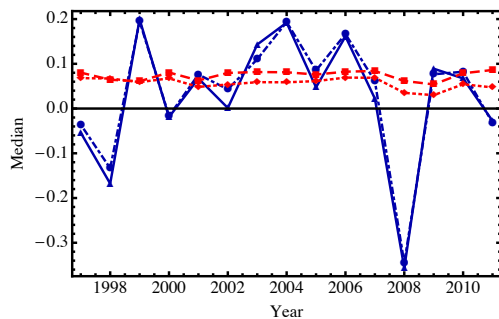
Figure 23: Time evolution of the median profit rate and the median growth rate of market value for India, Indonesia, Ireland, Israel, Italy, and Malaysia. Results are shown for the entire sample and the long-lived firms. For visual clarity, linear interpolations between annual data points have been added.



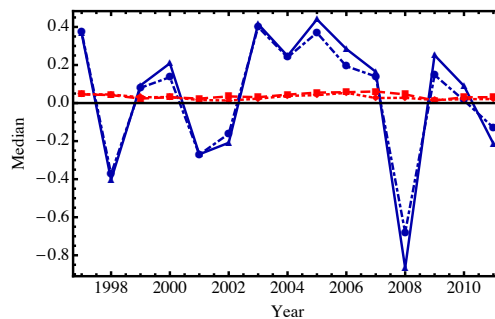
(a) Mexico



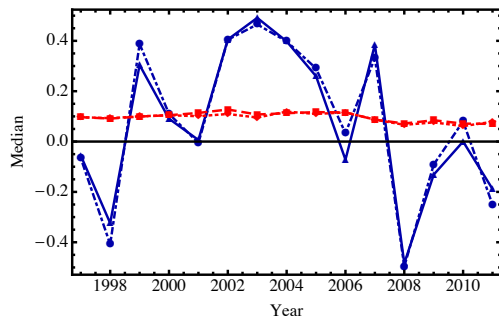
(b) Netherlands



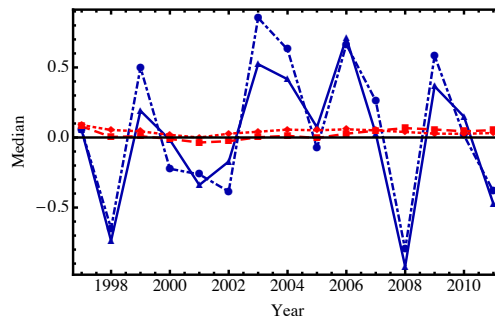
(c) New Zealand



(d) Norway



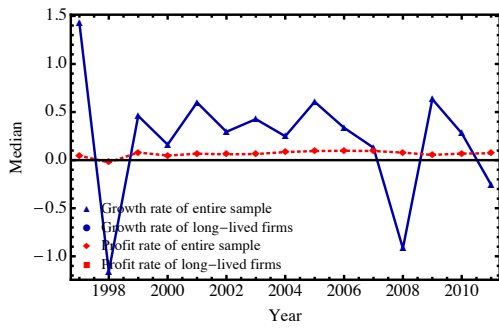
(e) Pakistan



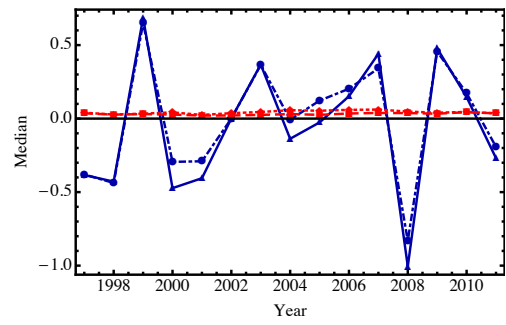
(f) Poland

Figure 24: Time evolution of the median profit rate and the median growth rate of market value for Mexico, the Netherlands, New Zealand, Norway, Pakistan, and Poland. Results are shown for the entire sample and the long-lived firms. For visual clarity, linear interpolations between annual data points have been added.

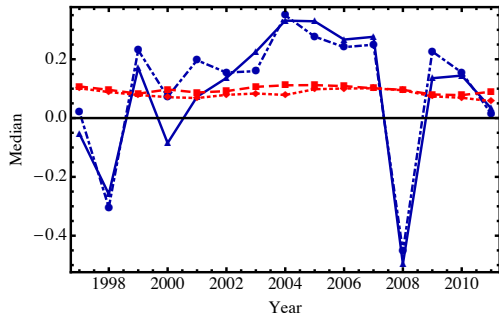




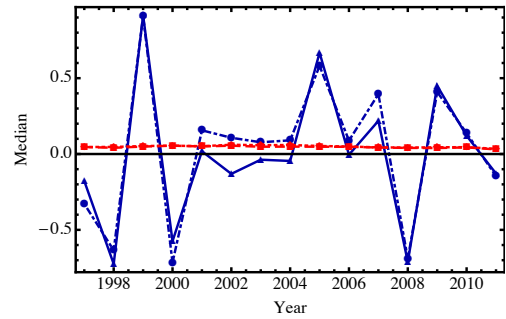
(a) Russia



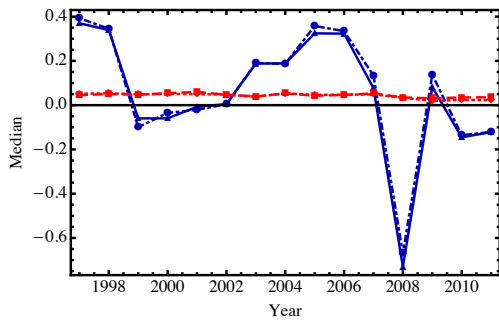
(b) Singapore



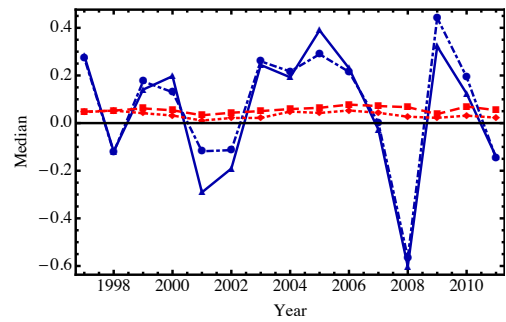
(c) South Africa



(d) South Korea

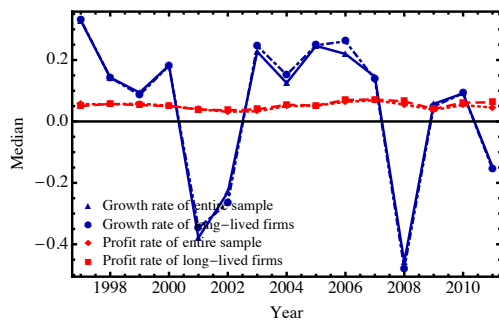


(e) Spain

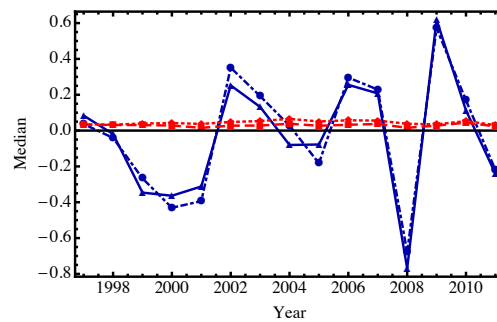


(f) Sweden

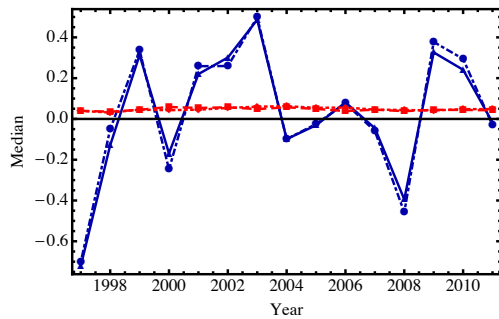
Figure 25: Time evolution of the median profit rate and the median growth rate of market value for Russia, Singapore, South Africa, South Korea, Spain, and Sweden. Results are shown for the entire sample and the long-lived firms. For visual clarity, linear interpolations between annual data points have been added.



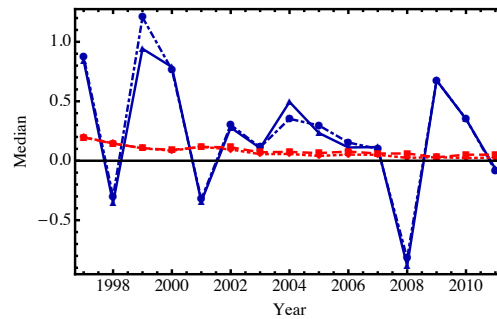
(a) Switzerland



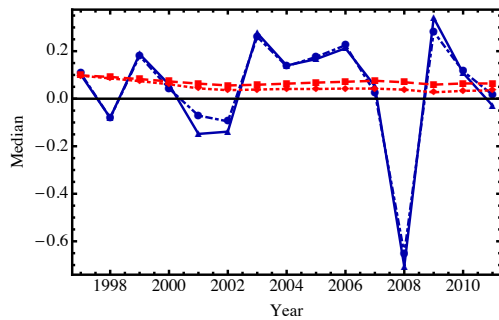
(b) Taiwan



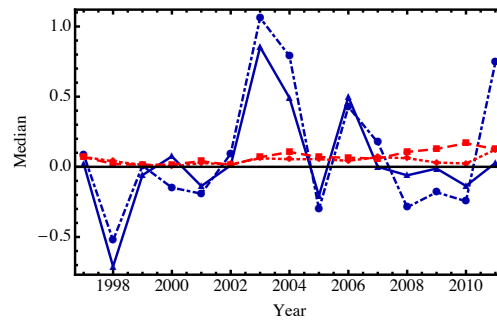
(c) Thailand



(d) Turkey



(e) United Kingdom



(f) Venezuela

Figure 26: Time evolution of the median profit rate and the median growth rate of market value for Switzerland, Taiwan, Thailand, Turkey, United Kingdom, and Venezuela. Results are shown for the entire sample and the long-lived firms. For visual clarity, linear interpolations between annual data points have been added.

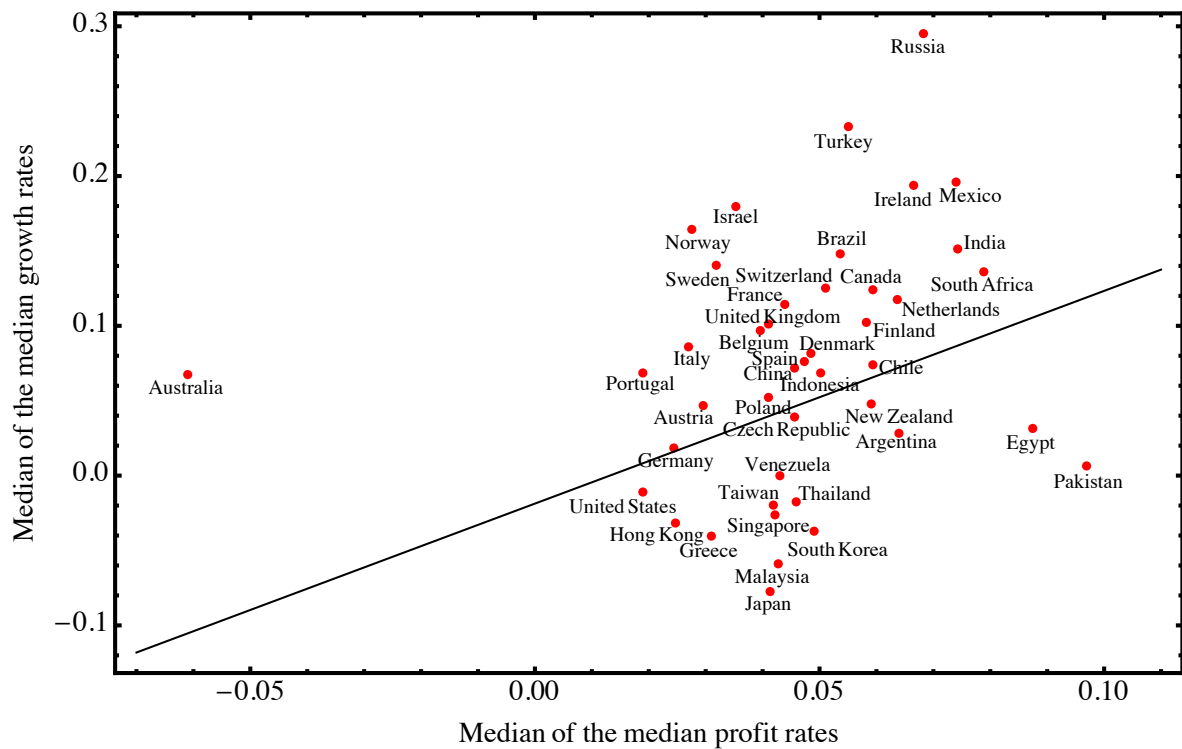


Figure 27: Scatter plot showing combinations of the average profit rate and the average growth rate of market value for the entire sample of firms. The data points have been computed as the median of the median time series. Weighted least squares regression of the average financial return on the average profit rate yields an intercept of  $-0.02 \pm 0.02$  with a p-value of 0.38 and a slope parameter of  $1.42 \pm 0.54$  with a p-value of 0.01. Weights have been calculated as the sum of market values of all firms in a country as a percentage of the market capitalization of the entire sample for a given year and are averaged over the period 1997-2011.

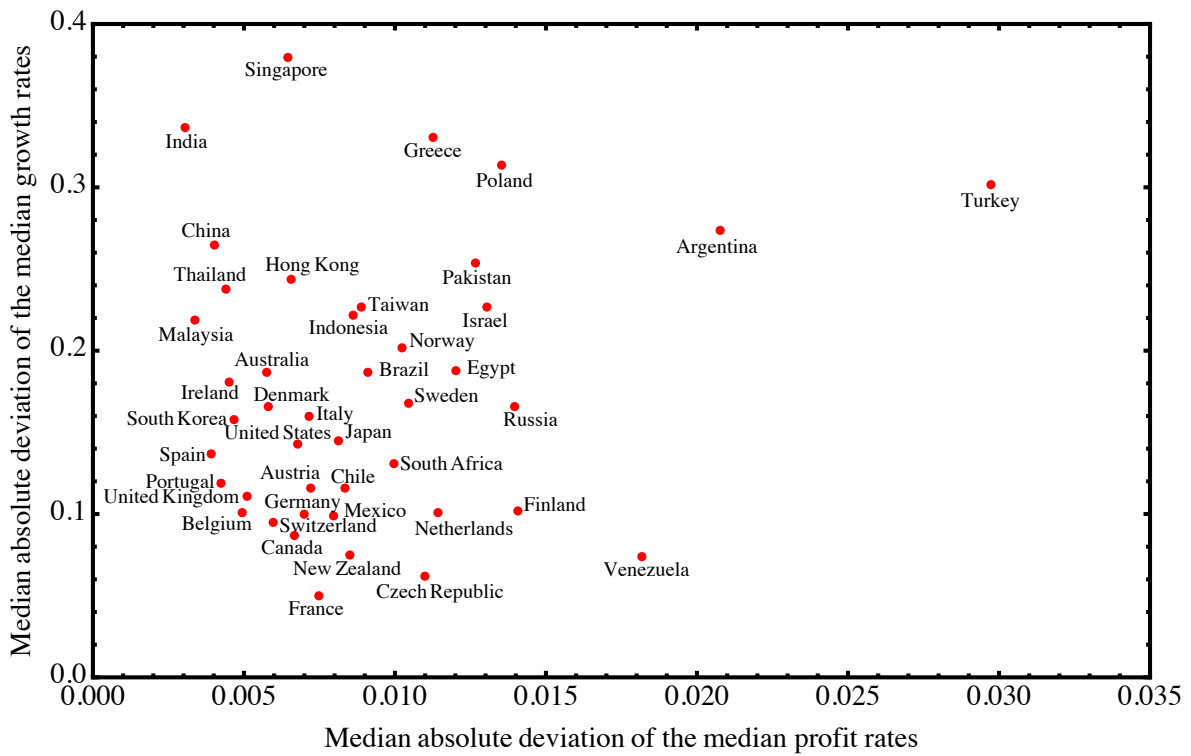


Figure 28: Scatter plot showing combinations of the profit rate volatility and the volatility of growth rates of market value for the entire sample of firms. The data points have been computed as the median absolute deviation of the median time series.

## THE ULTIMATE CORPORATE OBJECTIVE IS SURVIVAL

---

## ABSTRACT

The question about the determinants of firm profitability occupies research in industrial organization, strategic management, accounting and finance, and related fields. In one way or another, existing contributions suggest that either industrial or firm-level determinants of firm performance do exist. In this paper, I show that the effect of firm idiosyncrasies on the dynamics of profitability becomes almost negligible when companies survive in the market for a sufficiently long time period. Conditional on survival, the time evolution of the profit rate is governed by a common dynamic law for all firms. This has unexpected consequences for our understanding of corporate strategy because survival rather than profitability becomes the ultimate business objective.

“[...] Businessmen [...] tend to believe that generating profits is their principal mission, with survival and risk control something to perhaps consider - they miss the strong logical precedence of survival over success. To make profits [...], it would be a good idea to, first, survive.” (Taleb, 2012, p. 160)

### 3.1 INTRODUCTION

The return on capital is certainly one of the most important metrics to assess the performance of a business. It measures how much profit the entity generates per unit of its assets, and therefore allows to compare performance across different uses, e.g. firms or business units. The question about the determinants of this profitability ratio occupies practitioners as well as the academic profession, and lead to a plethora of theories and recommendations for business practice. In this paper, I study the diffusion of profitability for a diverse sample of long-lived corporations and look for systematic differences in its parametrization among firms. While previous studies stress the impact of firm and industry-level determinants on profitability, my results suggest that the profit rates of *surviving companies* are dispersed around a common measure of central tendency and have the same volatility on average, irrespective of the firms' individual characteristics. The perception that the profit rates of surviving corporations are realizations of the same stochastic process with a common location parameter and volatility for all firms has unexpected consequences for our understanding of corporate strategy because, in the long-run average, no surviving firm can do better, or worse, than the system-wide average. Consequently, survival becomes the primary business objective, concisely summarized in the introductory quote from Taleb (2012).

While the results reported in this paper suggest universality in the process governing profitability of corporations, the majority of existing investigations in industrial organization, strategic management, and accounting and finance stresses the sources of variation in profitability among firms (an excellent review of this literature, part of which is cited here, is provided in Goddard et al., 2005), putting different weights on

industrial, corporate, and business unit effects (see, e.g., Rumelt, 1982; Schmalensee, 1985, for a discussion of the relative importance of these effects). The industrial organization literature stresses the role of sectoral specificities. A prominent strand of this literature revolves around the well-known structure-conduct-performance paradigm which dates back to Mason (1939, 1949), Bain (1951, 1956), Porter (1980), and, more recently, Slater and Olson (2002). According to this approach, the structural characteristics of the market environment, e.g. the number and size of buyers and sellers, barriers to market entry, or firm diversification, affect the conduct of firms operating in this market which in turn determines their performance. The strategic management literature, on the other hand, puts more emphasis on the role of firm-specific effects (see, e.g., Barney, 1991, 2001; Levinthal, 1995; Peteraf, 1993; Teece, 1981). Proponents of the resource-based view, for example, argue that a sustained competitive advantage may arise from organizational, financial, technological, or intellectual resources available to the company which must be valuable, rare, and imperfectly imitable by others. Complementary insights stem from studies in accounting and finance that investigate the time series properties of performance measures, such as earnings or book rates of return to capital, to improve forecasts in equity valuation or cost of capital models. While early contributions in this vein conclude that earnings are essentially unpredictable and follow a random walk or submartingale process (Albrecht et al., 1977; Callen et al., 1993; Lintner and Glauber, 1978; Watts and Leftwich, 1977), more recent studies consider relative performance measures and stress the persistence in profitability (see, e.g., Fairfield et al., 2009; Fama and French, 2000; Freeman et al., 1982; Nissim and Penman, 2001). Like in industrial organization and strategic management, inter-firm differences in persistence and variability of profitability are associated with factors such as product type, barriers to market entry, firm size, and capital intensity (see, e.g., Baginski et al., 1999; Lev, 1983).

In this article, I consider some of the most popular quantities whose effect on profitability has been stressed in previous work, summarized in the extensive literature review in Table 13 in Appendix C.1, and study their impact on profitability over the lifespan of firms. I find that the relation between profitability and these characteristics is a highly time-dependent phenomenon: while average profitability and its volatility are remarkably homogeneous across long-lived firms and largely independent of firm characteristics, they exhibit (partly non-monotonic) correlations with these characteristics for shorter lived firms. The former result testifies to the existence of a common dynamic law governing the dynamics of surviving corporations that is obviously not intended by individual agents but results from the complex interactions among competitive firms. Given that the profit rates of surviving firms are drawn from the same distribution, their time series can be conveniently described by one and the same diffusion process with a common location and scale parameter for all companies. My findings are, therefore, in line with the notion of a statistical equilibrium which predicts that individual firm destinies have no systematic impact on the aggregate distributional outcome and the macroscopic properties of the system, summarized by the average profit rate and its volatility, which seem to be binding constraints for all surviving firms in the sample. Since these regularities hinge on the length of the ob-

ervation period, however, it is hardly surprising that studies which do not consider the long-run average behavior, but try avoid any notion of “survivorship bias”, yield mixed and inconclusive results on the sign and significance of profitability determinants. My results do not rule out the possibility that the factors identified in these and related studies may have an impact on the short to mid-run dynamics of profitability. Instead, the main argument here is that firm idiosyncrasies become negligible if firms survive in the market for a sufficiently long time period because the long-time average behavior of each firm converges to the average of the ensemble. The only variable that relates to the location and dispersion of profitability is the market to book value, or Tobin’s  $q$ , pointing towards a kind of dichotomy between “real” and “financial” firm attributes and a probably different perception of firm dynamics in financial markets and the real economy.

The notion of a statistical equilibrium in firm profit rates has been already stressed in previous investigations. Alfarano and Milaković (2008) encode the complex movements of capital in search of profit rate equalization into a moment constraint to derive the maximum entropy distribution of the rate of profit. This stationary distribution characterizes the cross-sectional distribution of states in equilibrium and is given by the Laplace which is parsimoniously described by a location and a dispersion parameter. A diffusion process for the time evolution of profitability that has the Laplace distribution as stationary density has been introduced in Alfarano et al. (2012). This process has three parameters: a location parameter, a dispersion parameter, and a diffusion coefficient. According to the statistical equilibrium methodology, the former two parameters should be identical across all firms and equal to the phenomenological values of the equilibrium distribution, while the latter parameter remains the only source of firm idiosyncrasies. The results presented in this paper are consistent with these implications of the model.

The structure of this article is as follows. After providing a brief summary of the statistical equilibrium model in Section 3.2, I introduce the data in Section 3.3. In Section 3.4, I estimate the diffusion process and investigate the relevance of firm idiosyncrasies for the diffusion of profit rates, considering groups of firms that differ with respect to their lifespan. The final section summarizes my findings and concludes.

### 3.2 MODEL REVIEW

The theoretical backbone of this investigation is the statistical equilibrium model of competitive firms that was presented in Alfarano et al. (2012). It builds on the classic idea that market economies are driven by the perpetual reallocation of capital in search of profit rate equalization. The starting point of the model is a distributional regularity in firm profit rates from which the authors derive a dynamic law of motion for profitability. This section provides a brief review of this statistical regularity and the model. Yet I refer the interested reader also to the initial paper for a detailed derivation and deeper discussion of the model.



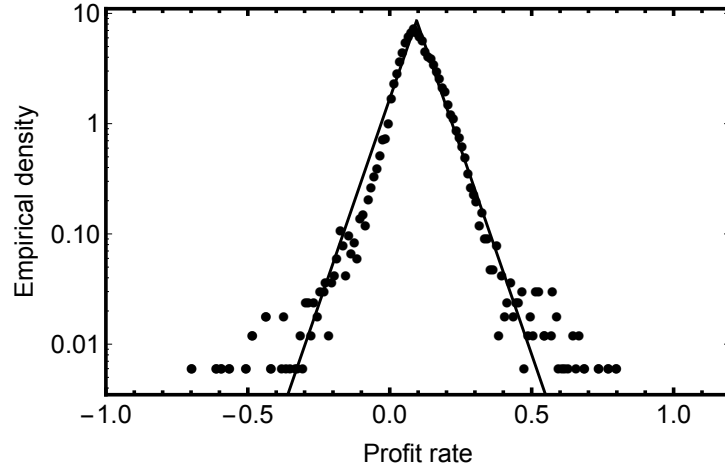


Figure 29: Pooled empirical density of annual profit rates for long-lived publicly traded US companies in the Datastream Worldscope database. Annual profit rates of all years between 1980 and 2012 are pooled together. The black line represents a fit of the Laplace distribution in equation (25) to the empirical density. Maximum likelihood estimation of the Laplace parameters yields  $\hat{m} = 0.0947 \pm 0.0006$  and  $\hat{\sigma} = 0.0581 \pm 0.0005$ .

Alfarano et al. (2012) start from the empirical finding that the *cross-sectional* distribution of long-lived firms' profit rates is not Gaussian but rather follows a symmetric Laplace (or double-exponential) distribution with probability density function<sup>28</sup>

$$f(x; m, \sigma) = \frac{1}{2\sigma} \exp\left(-\left|\frac{x - m}{\sigma}\right|\right). \quad (25)$$

The Laplace distribution in equation (25) is characterized by two parameters: a location parameter  $m \in \mathbb{R}$  and a dispersion (or scale) parameter  $\sigma \in \mathbb{R}^+$ . In Figure 29 I fit the Laplace to the cross-sectional profit rate density for the present sample of long-lived US corporations from the Datastream Worldscope database.<sup>29</sup> When plotted on semi-logarithmic scale, the data display the linear tent-shape that is characteristic of the Laplace distribution. Parametric estimation of the location and scale parameter by maximum likelihood yields  $\hat{m} = 0.0947 \pm 0.0006$  and  $\hat{\sigma} = 0.0581 \pm 0.0005$ , implying that surviving US companies generate an average profit rate of about 9.5 percent, along with a volatility of about 5.8 percent on average. In the further course of this study I will refer to these estimates as the phenomenological values of  $m$  and  $\sigma$ .

<sup>28</sup> Fat tails and deviations from Gaussianity imply the presence of interactions among firms and would not be observed if shocks to profitability were iid (Dosi et al., 2000). Thus, I presume that these statistical properties are due to some underlying correlating mechanism, first and foremost arising from the very process of competition.

<sup>29</sup> Details on the data will be given later in Section 3.3.

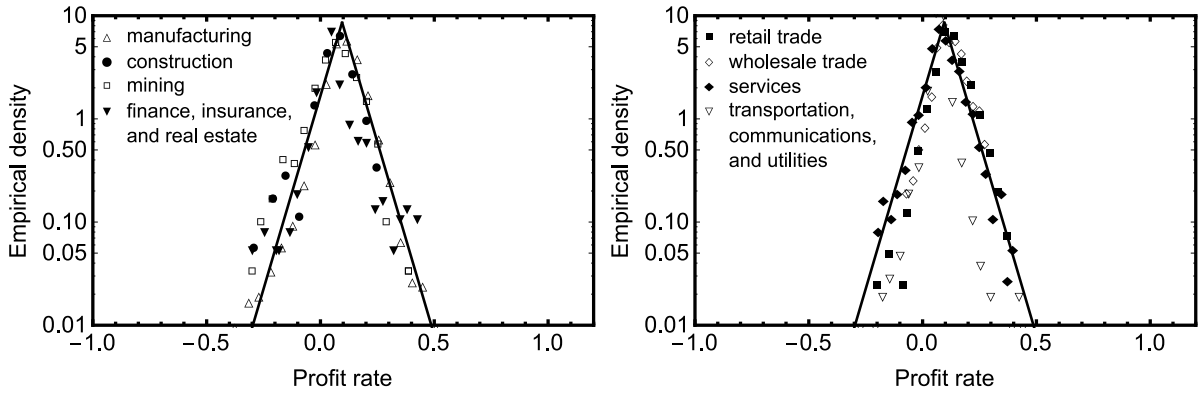


Figure 30: Sectoral distributions of annual profit rates for long-lived publicly traded US companies. The black lines illustrate a Laplace fit with the phenomenological values of  $m$  and  $\sigma$ .

The same distributional regularity is also observed on the sectoral level, as illustrated in Figure 30. To this end, I use two-digit SIC codes to group the companies into business divisions which represent the broadest categories according to the SIC scheme.<sup>30</sup> Then I plot the empirical density of their profit rates as in Figure 29, but this time for only one division at a time. Table 7 summarizes the corresponding parameter estimates. It is visually apparent that the profit rate distributions for particular sectors also exhibit the characteristic Laplacian shape and virtually collapse on the aggregate distribution, which suggests that this regularity is not an artifact of aggregation and testifies to the hypothesis that the fundamental force of profit rate equalization is an utterly universal mechanism that is present for all industries under study.

A stochastic process for the time evolution of the profit rate that is consistent with the Laplace distribution in equation (25) is the diffusion

$$dX_{i,t} = -\frac{D_i}{2\sigma_i} \text{sign}(X_{i,t} - m_i) dt + \sqrt{D_i} dW_t, \quad (26)$$

where  $dX_{i,t}$  is the change of firm  $i$ 's profit rate from period  $t$  to  $t + 1$ ,  $m_i$  is the average profit rate,  $\sigma_i$  is the volatility,  $D_i$  denotes the diffusion coefficient or characteristic time scale of the process,  $dW_t$  are Wiener increments, and  $\text{sign}(\cdot)$  is the signum function. The drift function in the first term of equation (26) captures the negative feedback mechanism that is described by the notion of classical competition: whenever the profit rate is below (above)  $m_i$ , the drift is positive (negative) and pulls the profit rate towards its long-term mean. The diffusion function in the second term of equation (26) is governed by the Brownian motion and reflects idiosyncratic factors on firm profitability. Since the latter are too complex for deterministic modeling, they are considered random.

<sup>30</sup> The division "agriculture, forestry and fishing" (SIC 01-09) is omitted because it contains only one long-lived firm. Details on the sectoral composition of the present sample are given later in Section 3.3.

The notion of statistical equilibrium is reflected in the parametrization of the dynamic law in equation (26). If the system is in statistical equilibrium, time series and cross-sectional averages coincide, implying that the parametrization of the dynamic law in equation (26) must be consistent with the *cross-sectional* distribution in equation (25). Formally, this requires the additional restriction  $m_i = m$  and  $\sigma_i = \sigma \forall i$ , i.e. all surviving companies face the same profitability benchmark and volatility on average, while all potential idiosyncrasies are captured by the diffusion coefficient  $D_i$ . The latter remains the only source of firm specific effects in the model because the equilibrium distribution is independent of the diffusion coefficient. In the following, I will turn to the empirical validity of this assumption. Before turning to the results, however, I will introduce the data set in the next section.

### 3.3 DATA

The sample of firms used in this study consists of publicly traded US enterprises reported in the *Datastream Worldscope* database between 1980 and 2012. The raw data set includes approximately 7,800 companies operating in 77 different sectors on a two-digit SIC code level. It is filtered according to the following criteria. First, I exclude banks (SIC codes 60 and 61) because the balance sheets of banking firms differ structurally from those in other industries. Second, I focus on long-lived companies which are defined as firms that are present in the data set for the entire period 1980-2012. As reported in Table 8, this filtering reduces the sample to 498 companies. Although these corporations represent less than 7 percent of the overall population in the sample, I would like to argue that they are a very interesting group of firms to study because they are mostly large firms which contribute disproportionately to aggregate output. In fact, Mundt et al. (2016) show that on average about 70 percent of these surviving firms belong to the power law tail of the firm size distribution (see Axtell, 2001). My perspective is therefore in line with a granular view on macroeconomic activity stating that microeconomic shocks to firms or disaggregated sectors do not cancel out in the process of aggregation as initially proposed by Lucas (1977), but may lead to significant aggregate fluctuations when the firm size distribution is sufficiently heavy-tailed (Gabaix, 2011), or in case of intersectoral input-output linkages (Acemoglu et al., 2012).

For the remaining 498 long-lived firms I obtain annual observations for the variables operating income, sales, property plant and equipment, current assets, total assets, common equity, total liabilities, long-term debt, short-term debt and current portion of long-term debt, market value of equity as well as the number of employees from Datastream. From the same source I also take a decomposition of total sales into up to ten product segments, sorted from the largest to the smallest segment according to its contribution to total sales. These data are complemented by the companies' year of foundation obtained from the corporate websites. Using the time series of operating income and total assets, I compute the return on assets for each firm which serves as an approximation to the profit rate. From the remaining variables I compute eight key variables intended to capture different financial and industrial characteristics of

Table 7: Maximum likelihood estimates of the Laplace distribution for different business divisions.

Division	$\hat{\mu}$	$\hat{\sigma}$
B: Mining	0.0744 (0.0044)	0.0717 (0.0005)
C: Construction	0.0802 (0.0052)	0.0584 (0.0038)
D: Manufacturing	0.1109 (0.0008)	0.0580 (0.0006)
E: Transportation, communications, electric and gas	0.0786 (0.0008)	0.0270 (0.0009)
F: Wholesale trade	0.1227 (0.0034)	0.0492 (0.0017)
G: Retail trade	0.1261 (0.0028)	0.0518 (0.0014)
H: Finance, insurance, and real estate	0.0355 (0.0010)	0.0815 (0.0062)
I: Services	0.0866 (0.0019)	0.0572(0.0015)

*Note:* Division A: agriculture, forestry and fishing is omitted since there is only one long-lived firm in this sector. Bootstrapped standard errors are shown in parentheses.

Table 8: Definition of business divisions and number of long-lived firms in each division.

Division	SIC codes	Number of firms
A: Agriculture, forestry and fishing	01 – 09	1
B: Mining	10 – 14	20
C: Construction	15 – 17	9
D: Manufacturing	20 – 39	267
E: Transportation, communications, electric, and gas	40 – 49	81
F: Wholesale trade	50 – 51	18
G: Retail trade	52 – 59	32
H: Finance, insurance and real estate	62 – 67	32
I: Services	70 – 89	38
Total		498

*Note:* Firms that operate in more than one sector are classified according to the sector that generated the highest revenue in 2012.

Table 9: Summary statistics for the independent variables.

Variable	Mean	SD	Min	25%	Median	75%	Max
Z <sub>1</sub> : Capital intensity	0.57	0.82	0.00	0.14	0.25	0.54	8.04
Z <sub>2</sub> : Firm age	91.69	37.87	15.00	60.00	89.50	115.00	222.00
Z <sub>3</sub> : Specialization	0.71	0.20	0.26	0.55	0.69	0.89	1.00
Z <sub>4</sub> : Leverage	0.64	0.59	-0.19	0.25	0.52	0.87	5.56
Z <sub>5</sub> : Productivity	267.52	238.76	23.13	124.48	184.29	331.86	1951.97
Z <sub>6</sub> : Liquidity	34.92	197.71	0.71	7.29	13.40	29.40	4229.90
Z <sub>7</sub> : Growth in sales	0.07	0.04	-0.06	0.04	0.06	0.09	0.25
Z <sub>8</sub> : Tobin's q	1.16	0.61	0.08	0.81	0.99	1.34	6.18

Note: I report (from left to right): mean, standard deviation, minimum, first quartile, median, third quartile, and maximum. As explained in the main text, explanatory variables represent time series medians over the sample period 1980-2012.

Table 10: Spearman rank correlations between the explanatory variables.

Variable	Z <sub>1</sub>	Z <sub>2</sub>	Z <sub>3</sub>	Z <sub>4</sub>	Z <sub>5</sub>	Z <sub>6</sub>	Z <sub>7</sub>
Z <sub>1</sub> : Capital intensity							
Z <sub>2</sub> : Firm age	0.11 (0.01)						
Z <sub>3</sub> : Specialization	0.03 (0.52)	-0.20 (0.00)					
Z <sub>4</sub> : Leverage	0.41 (0.00)	0.15 (0.00)	-0.13 (0.00)				
Z <sub>5</sub> : Productivity	0.18 (0.00)	0.13 (0.00)	0.00 (0.92)	0.32 (0.00)			
Z <sub>6</sub> : Liquidity	-0.48 (0.00)	-0.18 (0.00)	0.08 (0.07)	-0.61 (0.00)	-0.33 (0.00)		
Z <sub>7</sub> : Growth in sales	-0.15 (0.00)	-0.33 (0.00)	0.08 (0.06)	-0.14 (0.00)	-0.04 (0.32)	0.23 (0.00)	
Z <sub>8</sub> : Tobin's q	-0.09 (0.05)	-0.07 (0.13)	0.01 (0.84)	-0.36 (0.00)	-0.20 (0.00)	0.23 (0.00)	0.35 (0.00)

Note: P-values (in parantheses) equal to 0.00 imply  $p < 5 \times 10^{-3}$ .

the firms which turned out to be prevalent in prior studies of profitability (Capon et al., 1990). I explain these variables in more detail below. If time series of the firm characteristics are available, I calculate realizations for each year between 1980 and 2012 and take the median observation in order to obtain a more robust measure which is independent of extraordinary effects in single years. Summary statistics for the regression variables are provided in Table 9. Table 10 suggests that correlations between the independent variables, although partially significantly different from zero, are still low enough in absolute terms to expect sufficient independent variation among the independent variables, refuting potential concerns about concurvity.<sup>31</sup>

1. I operationalize *capital intensity* as the ratio of fixed capital, measured in terms of property plant and equipment, to total sales. Property, plant and equipment

<sup>31</sup> Concurvity denotes the nonparametric analog of multicollinearity.

is net of depreciation and includes portions of both owned and rented/leased fixed assets.

2. *Firm age* is measured in terms of the number of years from the firm's year of foundation to the latest period I have data for, 2012.
3. The *specialization* ratio rests upon the product segment decomposition of sales and measures the percentage of the largest product segment in terms of total sales. These product segments are defined on a 4-digit SIC code level. Consequently, the specialization ratio is larger than zero and smaller or equal to unity. The value one indicates that the firm is completely specialized, while a lower value stands for a higher degree of corporate diversification.
4. The sum of long-term debt and short-term loans divided by common equity represents *leverage*. According to general business practice, time intervals shorter than twelve months are regarded as short-term, while items with maturity longer than one year are long-term. Nevertheless, many firms include current portions in their long-term debt, which requires a reclassification of these amounts. Thus, I measure the amount of short-term loans via the Datastream variable "short-term debt and current portion of long-term debt" which backs out the current portion from long-term debt. Unlike alternative definitions of leverage that divide debt by total assets, my quite narrow definition attempts to exclude items that may have little to do with financing, e.g. trade credit or assets held against pension liabilities (see Rajan and Zingales, 1995, for a discussion of different measures of leverage).
5. I approximate *labor productivity* with the ratio of total sales to the number of employees. In Datastream the variable number of employees refers to the average annual number of full and part time employees, but excludes seasonal employees. Part time employees are converted into full time equivalents in Worldscope by default.
6. The *current ratio* is a proxy for liquidity and is intended to measure the firm's ability to repay short-term creditors with liquid assets. It is computed as current assets divided by short-term debt and current portion of long-term debt. The variable current assets includes cash and equivalents, receivables, inventories and prepaid expenses.
7. *Firm growth rates* are computed as logarithmic time differences in total sales.
8. Finally, I consider the well-known *q* ratio introduced by Tobin (1969) which compares the market value of a firm to the reproduction cost of its assets. For the present analysis I use the approximation of *q* suggested by Chung and Pruitt (1994) because it represents a good compromise between precision, data requirements and computational effort. Approximate *q* is calculated as the ratio of equity at market value plus the book value of short and long-term debt to total

assets, assuming that the reproduction costs of assets are equal to book values and that market and book values of debt are identical.

### 3.4 RESULTS

#### 3.4.1 Parameter estimation

I estimate the parameters of the diffusion process from its corresponding transient density using the Fokker-Planck equation formalism. With initial condition  $f(x_i, |x_{i,0}, 0) = \delta(x_i - x_{i,0})$ , where  $\delta(\cdot)$  denotes Dirac's delta function, and for appropriate boundary conditions, the solution to the Fokker-Planck equation reads

$$f(x_i, t|x_{i,0}, 0) = \frac{1}{\sqrt{2D_i\pi t}} \cdot \exp\left(-\frac{(x_i - x_{i,0})^2}{2D_i t} - \frac{1}{2\sigma_i}(|x_i - m_i| - |x_{i,0} - m_i|) - \frac{D_i}{8\sigma_i^2} t\right) + \frac{1}{2\sigma_i} \exp\left(-\frac{1}{\sigma_i}|x_i - m_i|\right) \Phi\left(-\frac{|x_i - m_i| + |x_{i,0} - m_i| - (D_i t)/(2\sigma_i)}{\sqrt{D_i t}}\right), \quad (27)$$

with  $\Phi(\cdot)$  the cumulative distribution function of the standard normal distribution (Toda, 2012).<sup>32</sup> Given discrete observations of the random variate that follows the diffusion process and the closed-form solution to its transitory density, we can apply maximum likelihood techniques to estimate the parameters of the diffusion process. To this end, for each time series  $\{X_{i,t}, t \geq 0\}$ , I numerically maximize the log-likelihood

$$\log \mathcal{L}(D_i, \sigma_i, m_i) = \log f(x_{i,0}; \sigma_i, m_i) + \sum_{t=0}^{T-1} \log f(x_{i,t+1}|x_{i,t}; D_i, \sigma_i, m_i) \quad (28)$$

with respect to the parameters, where  $f(x_0)$  denotes the (unconditional) Laplace density of the initial state  $x_0$  defined in equation (25), and  $f(x_{t+1}|x_t; D, \sigma, m)$  is the solution to the transient density evaluated for each observation  $x_{t+1}$  conditional on  $x_t$ . Two peculiarities of the estimation procedure deserve further explanation. First, I estimate  $m$  and  $\sigma$  separately and fix the other one of the parameters at its phenomenological value to increase the stability of the estimation. Moreover, fixing  $m$  in the estimation of  $\sigma$  ensures that the dispersion of profit rates is measured with respect to the systemwide benchmark  $\hat{m}$  for all firms.<sup>33</sup> Second, maximization of the likelihood function is non-trivial due the singularity of equation (25) at  $x = m$ . To cope with this issue, I prefer search algorithms to gradient-based methods and take recourse to the

<sup>32</sup> Toda (2012) uses the diffusion process as a model of income dynamics. His initial formulation of the Fokker-Planck equation does not contain the location parameter.

<sup>33</sup> Notice that the phenomenological  $\hat{\sigma}$  measures the dispersion of profit rates with respect to the phenomenological  $\hat{m}$ .

Table 11: Summary statistics for the parameter estimates.

	Mean	Median	Mode	SD	IQR	VC	QC	Skew	Kurtosis
$m_i$	0.1003	0.0920	0.0802	0.0568	0.0613	0.5662	0.6658	1.7203	13.4373
$\sigma_i$	0.0788	0.0633	0.0396	0.0729	0.0526	0.9259	0.8310	7.8879	94.4568
$D_i$	0.0946	0.0012	0.0002	1.5210	0.0028	16.0798	2.3083	18.7957	372.847

*Note:* I report the mean, median, mode, standard deviation, interquartile range, variation coefficient, quartile coefficient, skewness, and kurtosis. The mode is estimated with the half-range method proposed by Bickel (2002) which iteratively identifies densest half ranges.

differential evolution algorithm (Storn and Price, 1997) which does not use derivative information when attempting to find the global optimum of the likelihood function.<sup>34</sup>

Estimation of the diffusion process for each firm  $i = 1, \dots, n$  yields  $n = 498$  triples of estimates  $\{\hat{m}_i, \hat{\sigma}_i, \hat{D}_i\}_{i=1}^n$  whose sampling distributions are illustrated in Figure 31. Table 11 shows the corresponding summary statistics. The histograms of the estimated location and scale parameters are relatively narrow and peaked close to the phenomenological values  $\hat{m} = 0.095$  and  $\hat{\sigma} = 0.058$ , while the distribution of the diffusion coefficient is extremely right-skewed. It exhibits the heaviest tails and the highest variation of the three parameters. I take this to imply that heterogeneity among firms is captured primarily by the diffusion coefficient, while location and dispersion are relatively homogeneous across firms. While visual inspection of the histograms may be useful to get a first impression of the possible range of estimates, it cannot provide information on how much of the observed variability is due to sampling error and how much is due to significant differences among firms. To this end, I move from graphical analysis and descriptive statistics to more formal testing and apply the maximum likelihood ratio test. Table 12 illustrates the results. In case of the dispersion parameter, the null hypothesis that the estimated parameter is statistically indistinguishable from the phenomenological value cannot be rejected in 445 out of 498 cases on the 5 percent level which corresponds to approximately 89 percent of firms in the sample. For the location parameter, the null is not rejected in 431 cases which is about 86 percent of firms. I take this to imply that time series estimates of average profitability and volatility are largely consistent with the phenomenological values from the aggregate profit rate distribution.

Next, I turn to the question if the parameters of the diffusion process relate to firm characteristics.

### 3.4.2 Do firm idiosyncrasies matter for profitability?

To this end, I compute the expected value  $E[Y|Z] = g(Z)$ , where  $Y$  stands for either one of the three parameters of the diffusion process, conditional on the companies' industrial and financial characteristics introduced in Section 3.3. The latter are contained in the  $p$ -dimensional vector of regressor variables  $Z = (Z_1, \dots, Z_p)'$ . Since the functional

<sup>34</sup> Numerical analyses on the properties of the estimators are provided in Appendix C.2.2.



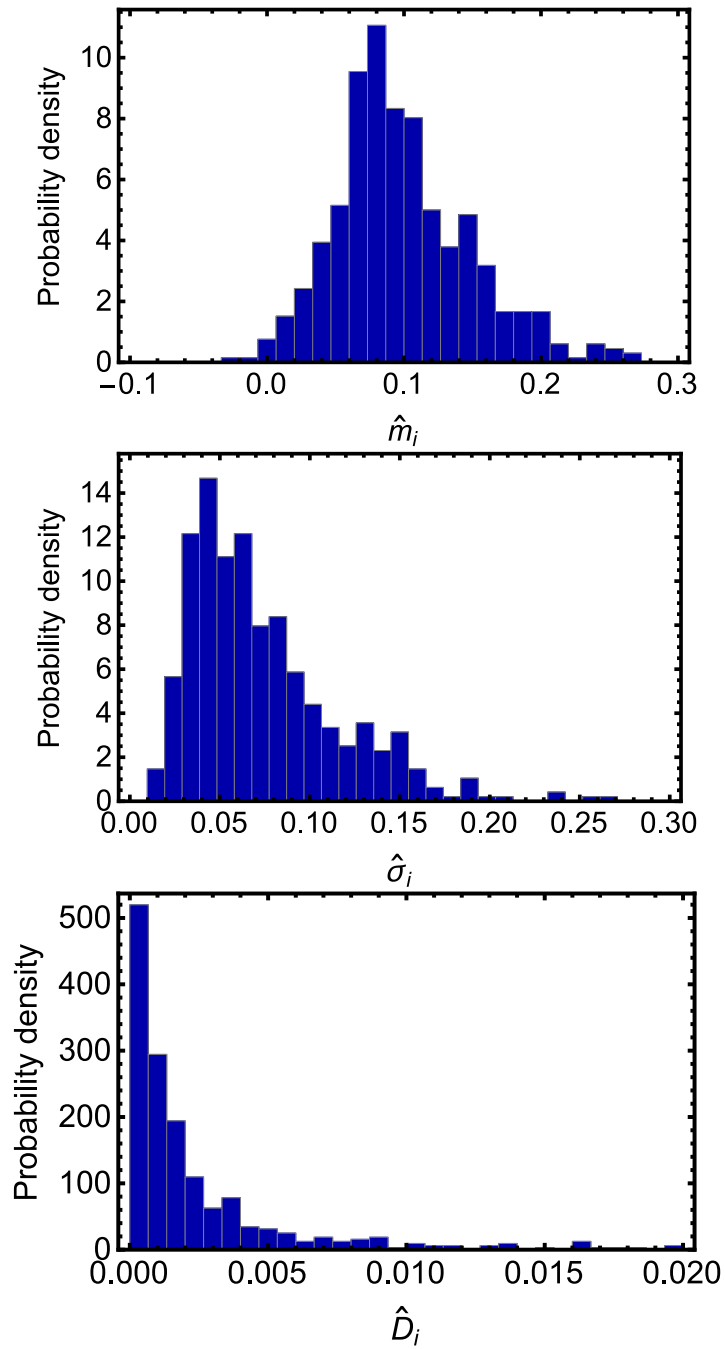


Figure 31: Histogram of location parameters, dispersion parameters, and diffusion coefficients.

Table 12: Likelihood ratio test.

	m	$\sigma$
Hypotheses:	$H_0: m_i = \hat{m}_i, (\sigma_i, D_i) = (\hat{\sigma}, \hat{D}_i);$ $H_1: m_i = \hat{m}_i, (\sigma_i, D_i) = (\hat{\sigma}, \hat{D}_i)$	$H_0: \sigma_i = \hat{\sigma}, (m_i, D_i) = (\hat{m}_i, \hat{D}_i);$ $H_1: \sigma_i = \hat{\sigma}_i, (m_i, D_i) = (\hat{m}_i, \hat{D}_i)$
Number of non-rejections of $H_0$ :	431	445

Note:  $\mathcal{L}_U$ : likelihood of the unrestricted model,  $\mathcal{L}_R$ : likelihood of the restricted model. Under  $H_0$  I impose the restriction that the parameter under test is equal to its phenomenological value, i.e. for the scale parameter I compare  $\mathcal{L}_U(\hat{\sigma}_i, \hat{m}_i, \hat{D}_i)$  to  $\mathcal{L}_R(\hat{\sigma}, \hat{m}_i, \hat{D}_i)$ , and for the location parameter  $\mathcal{L}_U(\hat{\sigma}, \hat{m}_i, \hat{D}_i)$  to  $\mathcal{L}_R(\hat{\sigma}, \hat{m}, \hat{D}_i)$ . The test statistic is  $\Lambda = -2\log\lambda$ , where  $\lambda = \mathcal{L}_R/\mathcal{L}_U$  denotes the likelihood ratio. For the scale parameter I compute p-values from the  $\chi^2$  distribution which is the asymptotic distribution of the test statistic under the null (Wilks, 1938). Since Wilks' theorem is violated in case of the location parameter, I use an approximation of the critical value from a simulated distribution generated via a resampling scheme. To this end, I simulate 10,000 realizations of the diffusion process with location and scale parameters equal to their phenomenological values, diffusion coefficient equal to  $D = 10^{-3}$  (which is the median coefficient computed over all long-lived firms), and time series length  $T = 33$ . Then I re-estimate the parameters with the procedure described in the main text, and compare the associated value of the likelihood function to the likelihood of the restricted model. The critical value is given by the 95% quantile of this simulated distribution (see Figure 37 in Appendix C.2.1).

form of these relationships is not obvious a-priori, I apply a nonparametric kernel regression technique that allows flexible estimation of a non-specific, smooth function  $g(\cdot)$  in

$$y = g(\mathbf{z}) + \epsilon, \quad (29)$$

where  $g$  is some unknown smooth function to estimate and  $\epsilon$  are observation errors. Extending this technique to the multiple regression case permits to control for potential interaction effects between the explanatory variables. The regression relationship in equation (29) is fitted with the multivariate generalization of the local constant estimator proposed by Nadaraya (1964) and Watson (1964)

$$\hat{g}_h(\mathbf{z}) = \frac{\sum_{i=1}^n \mathcal{K}_h(\mathbf{z} - \mathbf{Z}_i) y_i}{\sum_{i=1}^n \mathcal{K}_h(\mathbf{z} - \mathbf{Z}_i)} \quad (30)$$

(see e.g. Härdle et al., 2004).<sup>35</sup> It builds on the idea that observations  $\mathbf{Z}_i$  near the target point  $\mathbf{z}$  should contain information about the value of the regression function  $g$  at  $\mathbf{z}$  (Eubank, 1999). Thus,  $g$  is estimated as a weighted average of observed responses near the target point, with weights given by kernel functions  $\mathcal{K}$ . In the multiple

<sup>35</sup> All regressions are conducted in R using the *np* package by Hayfield and Racine (2008).

regression design,  $\mathcal{K}$  is a multivariate kernel function in the  $p$ -dimensional space which is constructed as a multiplicative kernel

$$\begin{aligned}\mathcal{K}_{\mathbf{h}}(\mathbf{z} - \mathbf{Z}_i) &= \frac{1}{h_1, \dots, h_p} \mathcal{K} \left( \frac{z_1 - Z_{i,1}}{h_1}, \dots, \frac{z_p - Z_{i,p}}{h_p} \right) \\ &= \prod_{j=1}^p h_j^{-1} K \left( \frac{z_j - Z_{i,j}}{h_j} \right),\end{aligned}\tag{31}$$

i.e.  $\mathcal{K}$  is the product of  $p$  univariate kernels  $K$ , which are continuous, bounded and symmetric real functions integrating to one.<sup>36</sup> Here, I choose one of the most commonly used functions and take  $K$  as a Gaussian kernel

$$K_{h_j}(z - Z_{i,j}) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left( \frac{z - Z_{i,j}}{h_j} \right)^2 \right\}.\tag{32}$$

The size of these kernel functions is determined by the parameter  $h_j$  which is called bandwidth.<sup>37</sup> Naturally, the bandwidths can differ across the different predictors. Thus, in the multivariate setting, we have a vector of bandwidths  $\mathbf{h} = (h_1, \dots, h_p)'$ . To fix the bandwidth parameter, I employ the multivariate generalization of the rule proposed by Silverman (1986)

$$h_j = \left( \frac{4}{p+2} \right)^{1/(p+4)} n^{-1/(p+4)} \psi_j,\tag{33}$$

where  $n$  denotes sample size, and  $\psi$  is a robust measure of spread (Scott, 2015).<sup>38</sup> Finally, I evaluate the uncertainty of the regression function using 50 pointwise confidence intervals based on bootstrapped standard errors. These are adjusted for multiple testing with the Bonferroni correction.<sup>39</sup>

Figure 32 illustrates the conditional estimates of the location parameter. Except for Tobin's  $q$ , the partial regression plots are almost flat for all variables, implying that profitability does not exhibit any systematic relationship to the industrial and organizational characteristics of the companies *on average*. In some cases we observe some

<sup>36</sup> The product kernel constructs the unknown multivariate density of all covariates from the product of their univariate kernels. The idea is that, under independence, the joint density of random variates can be computed as the product of their marginal densities. Notice, however, that the assumption of independence among the kernels does not transfer over to the multivariate estimator because averaging is still conducted over observations from all covariates (see Henderson and Parmeter, 2015, p. 61), i.e. the independence assumption is used to construct the kernel weights, but not the entire estimator.

<sup>37</sup> The bandwidth parameter determines the smoothness of  $\hat{g}$ . A small bandwidth leads to a relatively noisy regression surface, whereas a too large value results in oversmoothing and might eliminate important features of the relationship. Therefore, an appropriate value of the bandwidth parameter has to be selected that balances bias and variance of the estimate.

<sup>38</sup> To increase robustness,  $\psi$  is estimated as the minimum of the sample standard deviation and the normalized interquartile range,  $\hat{\psi}_j = \min \left\{ \frac{1}{n-1} \sum_{i=1}^n (z_{i,j} - \bar{z}_j)^2, \text{IQR}/1.34 \right\}$ .

<sup>39</sup> For visual convenience, I draw continuous bands around the 50 error bars in all graphical representations of the regression function.

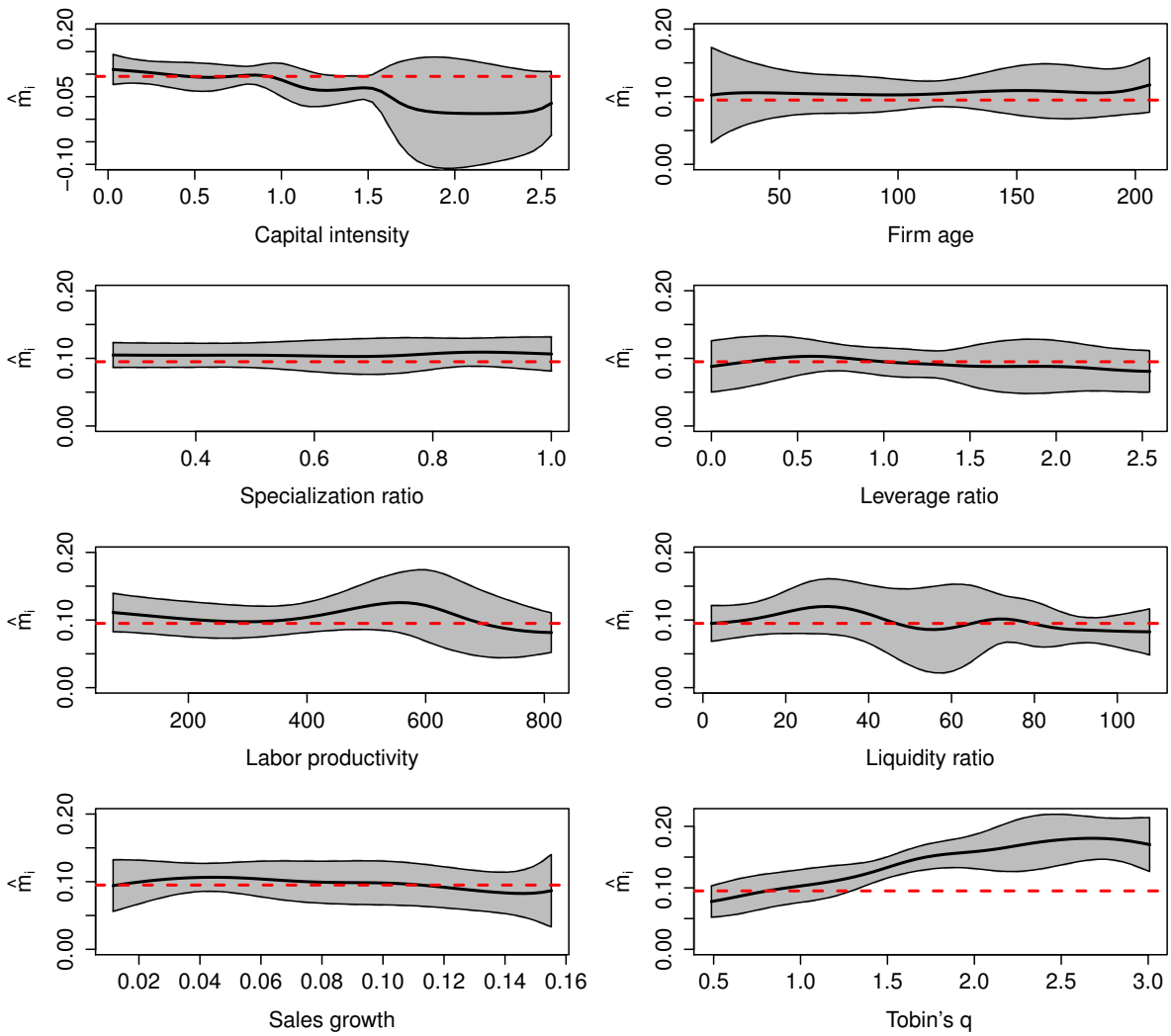


Figure 32: Partial nonparametric regression plots for the location parameter with 95% confidence intervals. The two-dimensional graphical representation of the multiple regression relationship is obtained by plotting the location parameter against one predictor and fixing all other explaining variables at their median value. Dashed horizontal lines are a visual aid and illustrate the phenomenological value  $\hat{m} = 0.0947$ . Since bandwidths are constant over the entire support of each variable and the data may become scarce in the tails of the distribution, the data are trimmed by dropping between one and three percent of the most extreme realizations at both ends of the spectrum.

mild fluctuations in the conditional means, yet the 95 percent confidence bands always enclose the phenomenological value  $\hat{m} = 0.095$ . The only statistically significant relationship is the market to book value which relates positively to the location parameter, i.e. companies that operate below the system-wide benchmark are valued less than the reproduction cost of their assets, while the securities market assigns a kind of premium to those firms which are excessively profitable. My findings, therefore, suggest a kind of dichotomy between “real” characteristics of the companies like growth, age, or diversification on the one hand, and the firms’ financial valuation on the stock market on the other hand. The relationship between the average profit rate and  $q$  may not be surprising considering that both quantities are measures of corporate performance. Unlike the profit rate, however,  $q$  should incorporate aggregate expectations about the stream of future profits via marking a firm’s assets to market (Erickson and Whited, 2000).<sup>40</sup> Thus, one could argue that the market-to-book ratio increases with average accounting returns because these raise expectations about future profitability (Pastor and Veronesi, 2003). However, considering that the diffusion process predicts mean-reversion of the return to capital, such deviations from the systemwide-average profit rate cannot be sustainable in the long-run and should be eliminated through the reallocation of capital. From this perspective, it strikes me that financial valuation somehow overreacts to the short and mid-run dynamics of profitability, without taking into account that a common location and scale parameter reflects the cost of survivorship.<sup>41</sup> To strengthen the argument that financial variables behave somewhat differently, I have also experimented with alternatives to  $q$ . For example, I considered the ratio of market value of equity to total assets as well as the ratio of market value to the book value of equity. It turns out that  $m$  also relates to these variables which confirms that quantities involving market valuation exhibit peculiar patterns.

Results for the scale parameter, displayed in Figure 33, are qualitatively in line with those obtained for the location parameter. Companies with higher  $q$  exhibit more volatility in the profit rate which should be attributed to the deviation of  $\hat{m}_i$  from  $\hat{m}$  for these firms.<sup>42</sup> For the remaining variables I do not detect any systematic effect on the dispersion of profit rates.

To assess the power of my results, I performed several robustness checks. First, it turns out that average profitability and its volatility remain independent of “real” characteristics if  $q$  is excluded in the regressions. Thus, I can rule out that independence of  $m$  and  $\sigma$  merely prevails because all volatility in these parameters is captured by  $q$ . Second, I used also shorter time intervals (15, 20, 25 years of data) to estimate the parameters for long-lived firms to see if independence is due to long-term averag-

<sup>40</sup> In investment theory, Erickson and Whited (2000) interpret the expected stream of future marginal benefits from investment as *marginal*  $q$ . However, since marginal  $q$  is difficult to measure, most empirical studies (including the present one) consider *average*  $q$  instead. Hayashi (1982) shows that the two measures coincide if firms are perfectly competitive and production technology exhibits constant returns to scale.

<sup>41</sup> The notion of a real vs. financial disconnect is, of course, well established in the literature, mostly revolving around excess volatility in financial markets (e.g. LeRoy and Porter, 1981; Livan et al., 2014; Mundt et al., 2014; Shiller, 1979, 1981).

<sup>42</sup> Again, notice that volatility is measured with respect to  $\hat{m}$  for all firms.

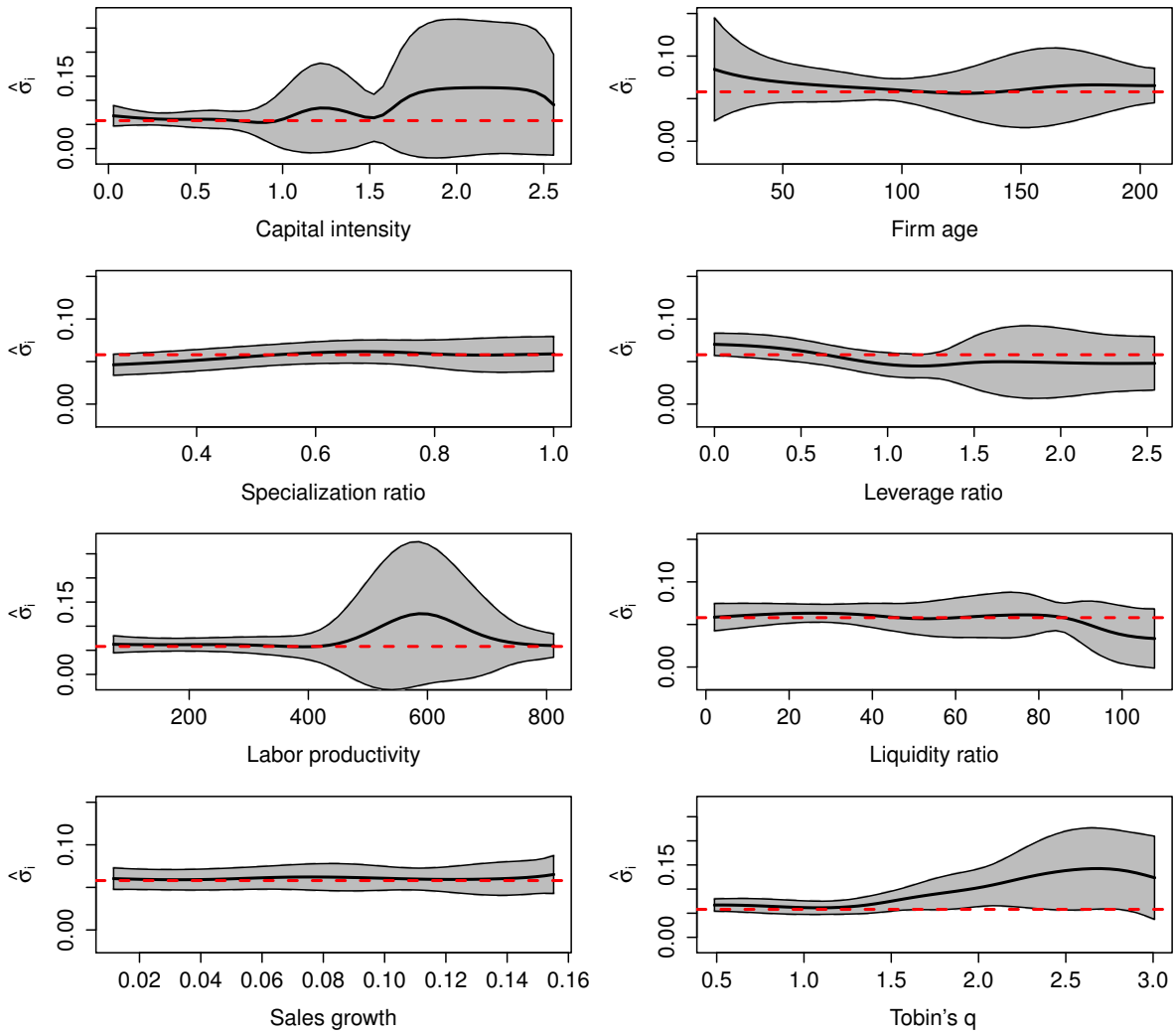


Figure 33: Partial nonparametric regression plots for the scale parameter with 95% confidence intervals. Horizontal lines are a visual aid and illustrate the phenomenological value  $\hat{\sigma} = 0.0581$ .

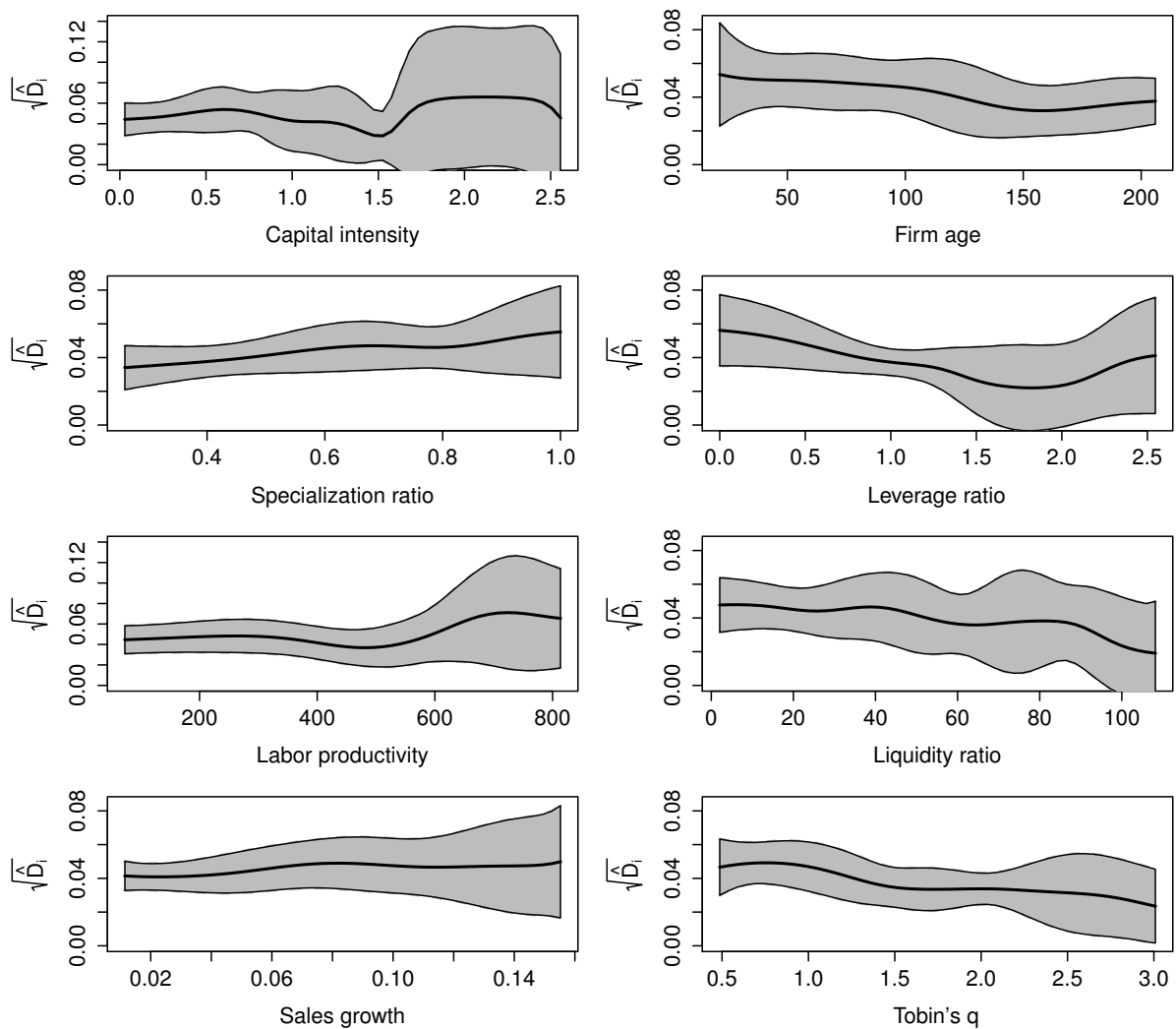


Figure 34: Partial nonparametric regression plots for the diffusion coefficient with 95% confidence intervals.

ing. It turns out, however, that the results are virtually the same for shorter training samples. Third, I investigated the relation between the two parameters and time series volatility of corporate characteristics to check if changes in these characteristics might explain differences in profitability. These regressions, graphically illustrated in Figures 38 and 39 in Appendix C.3, suggest that even the time volatility of firm attributes is irrelevant for average profitability and its volatility. I take this as another, markedly stronger, indication that location and dispersion are truly independent of idiosyncratic efforts and firm specificities.

Figure 34 shows the diffusion coefficient as a function of the eight explanatory variables. Contrary to the former two parameters, we now observe some (weak) dependencies on “real” characteristics, e.g. firm age, specialization, leverage, and liquidity, which is consistent with previous findings in the persistence of profit literature (e.g. Geroski and Jacquemin, 1988; Gschwandtner, 2005; Mueller, 1977, 1990; Waring, 1996).

The speed of adjustment is lower for mature, less specialized, and liquid firms. For some variables, e.g. leverage, the observed relationships exhibit even non-monotonic patterns. It is also worth noting that  $q$  is negatively correlated with the diffusion coefficient, implying that firms operating under a lower diffusion coefficient exhibit higher market-to-book ratios on average. Overall, the finding that  $D$  captures firm idiosyncrasies is consistent with the notion of statistical equilibrium because the diffusion coefficient does not determine the stationary equilibrium distribution in equation (25).

### 3.4.3 *Are shorter lived firms different?*

The results that have been presented so far suggest that average profitability and its volatility do not depend on industrial characteristics, while the diffusion coefficient captures firm idiosyncrasies. In this section, I investigate if this property also holds for shorter lived firms. Put differently, I study the effect of sample selection considering the conditional mean of the parameters as a function of survival time. Since my focus is on long-lived corporations while the majority of previous studies in the field attempts to avoid what is usually called “survivorship bias”, one could hypothesize that opposing views on the relevance of firm idiosyncrasies might result from the consideration of different samples and life spans. To this end, I group firms according to their survival time  $T$  into one of three disjunct subsamples. The first group contains 1804 firms which exist in the population surveyed by the database for 10-17 years. The second sample contains 837 firms which exist for 18-25 years, while the third group contains all older companies (720 firms).<sup>43</sup> Since it is well-known that the profit rates of shorter lived firms do not obey a symmetric Laplace but an even more leptokurtic and asymmetric distribution (Alfarano et al., 2012; Wagner et al., 2010), I do not estimate the location and dispersion parameter from the diffusion process for these firms. Instead, two descriptive statistics are considered. My estimate of the location parameter is the median profit rate of the time series, while volatility is measured in terms of mean absolute deviation from  $\hat{m}$ . Then, as before, I regress the parameters on the firm characteristics, but this time separately for each group of firms.

<sup>43</sup> I am aware that the number of observations is only a rough proxy for survival time because firms could have been incorporated before 1980, the first year of the observation period, implying that the true survival time is longer. Unfortunately, data on the year of incorporation is not available for all approximately 7,800 firms in the sample, but only for the 498 long-lived corporations, so that I cannot compute a more accurate measure of survival time at this point. To assess the significance of this shortcoming, I considered the firms with data in 1980 in more detail. It turns out that 543 firms are present in the sample in that year, 498 (i.e. more than 90 percent) of which are classified as long-lived corporations. 30 out of the remaining 45 firms exist in the sample for 25 years or more, i.e. these entities already belong to the group of firms with the longest survival time. 5 out of the remaining 15 corporations have a life span of 10-17 years, while the other 10 companies belong to the second group with  $10 \leq T \leq 25$ . Since the risk that firms have been incorrectly classified as shorter lived entities with  $10 \leq T \leq 17$  affects at most 5-15 firms, I would like to believe that my definition of life span is a reasonable approximation, and that the measurement error is still small enough to obtain meaningful results.



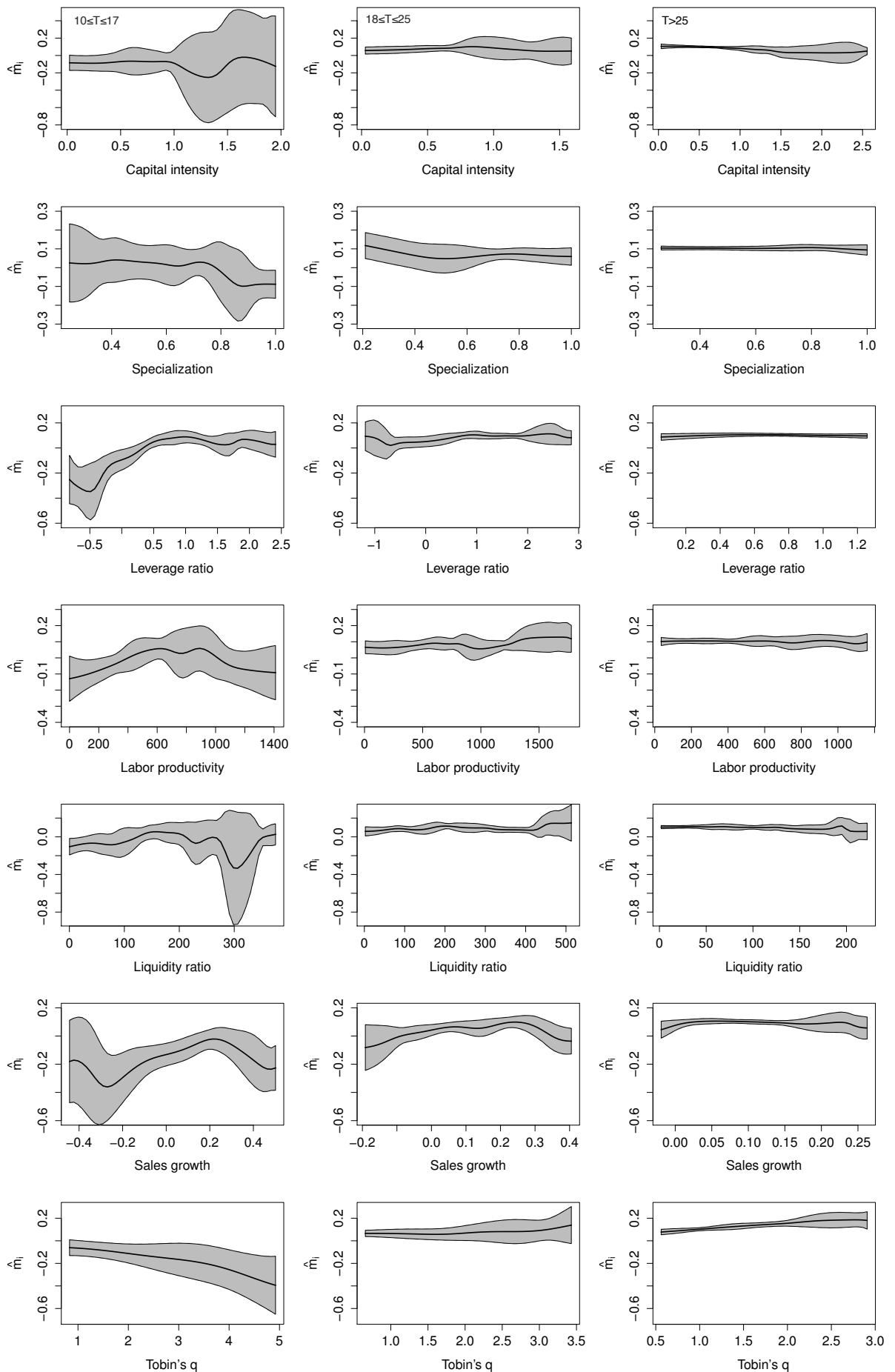


Figure 35: Partial nonparametric regression plots for the location parameter conditional on life span  $T$ . The right panel shows firms with 10-17, the middle panel those with 18-25, and the right panel those with more than 25 observations.

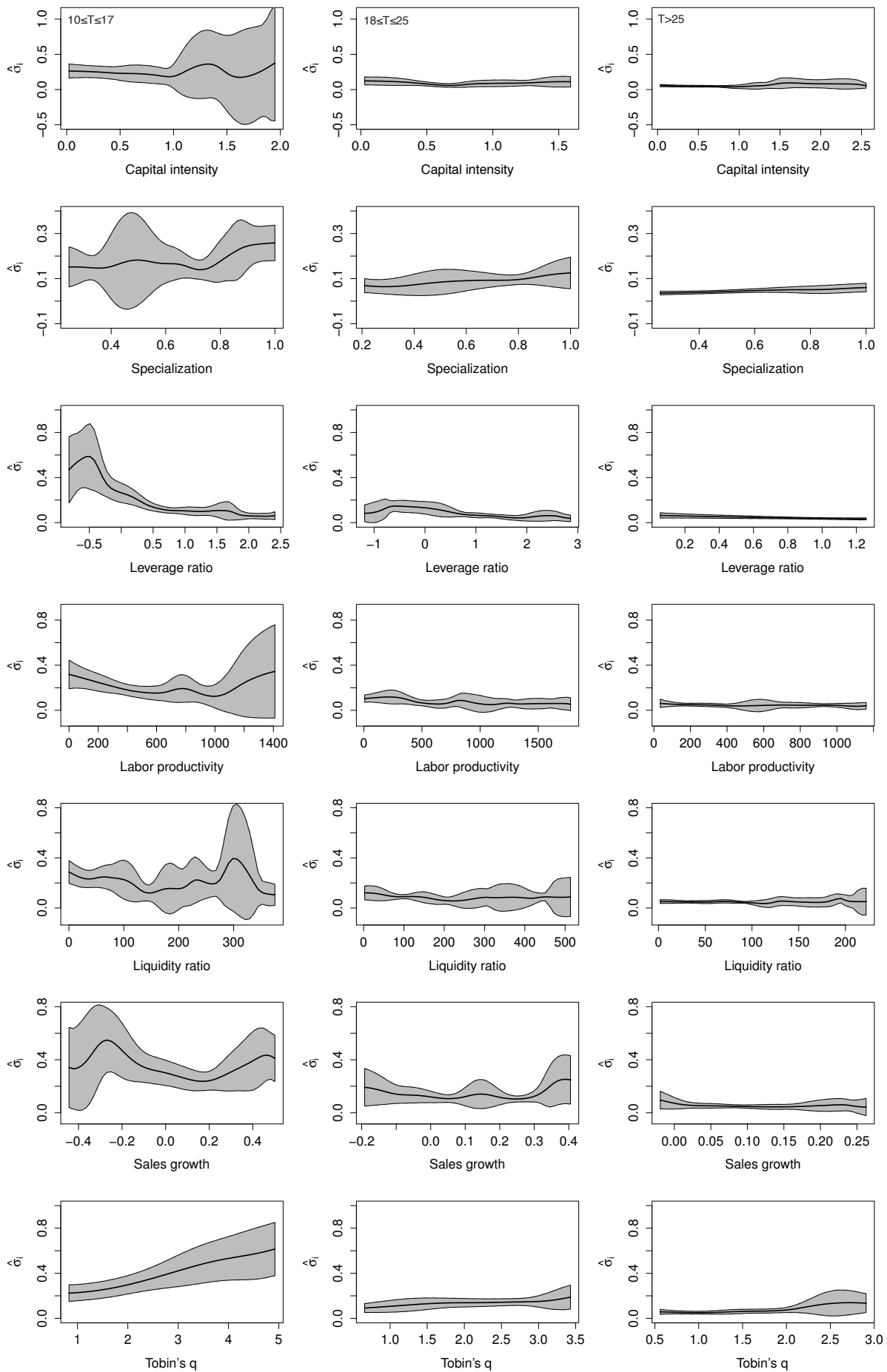


Figure 36: Partial nonparametric regression plots for the dispersion parameter conditional on life span  $T$ . The right panel shows firms with 10-17, the middle panel those with 18-25, and the right panel those with more than 25 observations.

Figures 35-36 illustrate the regression results for the three subsamples. It is visually apparent that the parameters for shorter lived firms relate to (at least some of) the characteristics of the firms under study. For  $10 \leq T \leq 17$ , correlations not only occur between profitability and  $q$ , but also pertain to other variables. Often, these relationships exhibit even non-monotonic patterns, which is consistent with the mixed and sometimes inconclusive results on these relationships that have been reported in previous studies (see Table 13 in Appendix C.1). If the focus shifts to long-lived corporations, however, firm idiosyncrasies become clearly less important. Moreover, comparing firms in the left panels of Figure 35 to those in the right panels, it is worth noting that the sign of the relationship between average profitability and  $q$  changes from negative to positive. At this point I can merely speculate, but one could argue that this speaks in favor of some sort of uncertainty effect. Along the lines of Pastor and Veronesi (2003), the market-to-book ratio increases with the uncertainty about future profitability. This uncertainty contributes to a high market valuation of young, unprofitable firms. Since the uncertainty declines over time as investors learn about the posterior profitability distribution, one should expect a gradual decline of the market-to-book ratio over time. At the same time, however, there is a positive effect of increasing profitability on the market value via the expectations channel. So, while the uncertainty is seemingly dominating for shorter lived firms, the expectations channel is more relevant for long-lived companies. The notion of increasing (average) profitability and declining volatility over the lifespan of firms is confirmed in Figure 40 in Appendix C.4.

### 3.5 CONCLUDING REMARKS

This paper investigates the impact of firm idiosyncrasies on the dynamics of corporate profitability. In contrast to previous studies, which stress the importance of firm and industry-level determinants (e.g. diversification, productivity, capital structure), I find that the profit rates of surviving corporations are on average the same and also exhibit equally volatile fluctuations, irrespective of most of their idiosyncratic characteristics. Tobin's  $q$  (and other variables that contain the market valuation of corporations), on the other hand, correlate significantly with profitability and its volatility, pointing towards a dichotomy between real and financial attributes. This disconnect is consistent with the notion of bubbles and fads (Blanchard et al., 1993), or any other social interactions that lead to herding behavior in financial markets, that is changes in financial valuation which cannot be justified ex post by fundamentals (see, e.g., the seminal story by Shiller, 1981, on subsequent dividend or earnings changes). Interestingly, Livan et al., 2014 have recently argued that the diffusion model provides an alternative, yet constructive, measure of the real performance of corporations that does not rely on dividends or earnings. Their main idea is to compare profit rate cross-correlations with financial cross-correlations using a spectral decomposition of the respective time series. The problem is that profit rates are positively correlated in time, while financial returns are not (in line with the efficient market hypothesis by Fama, 1970, 1991). The diffusion model allows them to filter out the serial correlation in real returns, and

they are able to show that merely subtracting the leading spectral component from the financial correlations (often referred to as the market mode or excess comovement) makes the two time series consistent with each other, both in terms of the average return and the average volatility. This further testifies to the usefulness of the process, even in light of the real vs. financial disconnect. Certainly, the positive correlation between profitability and financial performance does not inform us about the causality between the two. At any rate, it is hard to imagine that financial performance would have a causal influence on real performance, while it seems intuitive that the opposite has to be the case. It is worth noting, however, that there is a recent strand of literature that tries to make exactly the former case (see, e.g., Goldstein and Yang, 2014).

One of the most significant findings of this study is that a common law for corporate profitability prevails only for surviving entities, i.e. firms that managed to withstand the competitive forces over an extended period of time, while this is not true for shorter lived firms. Thus, it is potentially the consideration of different samples that leads to these diverging views on the role of firm or industry-level characteristics in the existing literature and in the present study. While my results might seem unexpected or even contestable to authors who have been working in the many fields relating to corporate profitability, it should be noted that they are entirely data-driven. After all, Figures 35 and 36 strikingly show the independence of profitability and its volatility from corporate idiosyncrasies conditional on survival time by solely relying on descriptive statistics. This is not to say that the reduced-form model is peripheral to our investigation. Quite to the contrary, it is only by virtue of the model's accuracy that I was led to investigate the stark, and in light of the literature most unexpected, conjecture that idiosyncrasies are irrelevant for corporate profitability.

Finally, the finding that the profit rates of long-lived corporations appear as realizations of the same stochastic process, with a common location and dispersion parameter for all firms, has instantaneous consequences for our understanding of corporate strategy because, conditional on survival, corporations cannot do better (or, perhaps more surprisingly, worse) than the system-wide average. Consequently, survival ought to be the primary, if not the only, business objective - and is everything but just a short-term goal for small, young firms.

## APPENDIX

## C.1 LITERATURE REVIEW

Table 13: Determinants of profitability in the empirical and theoretical literature.

Determinant	Author(s)	Type	Sample	Period	Control variables	Relationship
Diversification	Montgomery (1985)	E	128 Fortune 500 firms	1972-1977	Market share, firm size, concentration, firm growth	Negative
	Palich et al. (2000)	M				Inverted-U-shaped
	Palepu (1985)	E	30 firms from the food products industry	1973-1979		Neutral
	Tallman and Li (1996)	E	192 large US multinational manufacturing firms	1987	Firm size, leverage, industry growth	Quadratic
	Markides and Williamson (1994)	E	164 randomly selected Fortune 500 firms	1981	Capital intensity, R&D intensity, advertising intensity	Positive

(continued)

Table 13 – Continued

Determinant	Author(s)	Type	Sample	Period	Control variables	Relationship
Market valuation	Blanchard et al. (1993)	T				Ambiguous: when asset markets value firms at their fundamental value, one should expect a tight positive relationship between the profit rate and Tobin's q. On the other hand, deviations may occur in case of information asymmetries between managers and the market, fads or speculative bubbles
	Pastor and Veronesi (2003)	T				Positive: the market-to-book ratio increases with average accounting returns because these raise expectations about future profitability. Moreover, the market-to-book ratio increases with uncertainty about future profitability
Growth	Reid (1995)	E	73 young firms	Scottish micro 1985-1988	Business type, market share, rival's pricing, product differentiation, leverage	Negative

(continued)

Table 13 – Continued

Determinant	Author(s)	Type	Sample	Period	Control variables	Relationship
	Cho and Pucik (2005)	E	488 Fortune 500 firms	1998-2000	Quality and innovativeness indices, market value	Positive
	Bottazzi et al. (2008)	E	Unbalanced panel of 14,000-17,000 Italian manufacturing and 10,000-13,000 service firms	1998-2003	Credit ratings, sectoral affiliation	Neutral
Leverage	Rajan and Zingales (1995)	E	Approximately 8,000 international corporations from 31 different countries	1987-1991	Firm size, market-to-book ratio, ratio of fixed to total assets	Negative
	Fama and French (2002)	E	Approximately 3,000 firms from the Compustat database, excluding utilities and financial firms	3 years between 1965 and 1999	Market-to-book ratio, depreciation-to-assets ratio, firm size, R&D expenditures relative to total assets	Negative
	Danis et al. (2014)	E	Non-financial firms from the Compustat database	1984-2011	Firm size, market-to-book ratio, tangible assets, industry concentration	Non-monotonic: positive when firms are at or close to optimum leverage, and negative otherwise
Liquidity	Goddard et al. (2005)	E	Pan-European sample consisting of 12,508 firms	1993-2001	Past profitability, firm size, market share, leverage	Positive
	Deloof (2003)	E	1,009 large Belgian firms	1992-1996	Firm size, firm growth, financial debt ratio, ratio of fixed financial to total assets	Positive

(continued)

Table 13 – Continued

Determinant	Author(s)	Type	Sample	Period	Control variables	Relationship
	García-Teruel and Martínez-Solano (2007)	E	8,872 small and medium-sized Spanish firms	1996-2002	Firm size, firm growth, leverage	Positive
	Eljelly (2004)	E	29 Saudi-Arabian publicly-traded firms	1996-2000	Firm size	Negative
Capital intensity	Russo and Fouts (1997)	E	243 US firms from various sectors, excluding utilities	1991-1992	Firm growth, advertising intensity, firm size, industry concentration, industry growth, environmental rating	Negative
	Hecht (2008)	E	Unbalanced panel of approximately 2,400 active and inactive nonfinancial firms located in Japan, Germany, UK and US	1980-2004	market-to-book ratio, ratio of dividends to book equity, country, industry, dividend payments	Negative
	Fama and French (2000)	E	Unbalanced panel of 2,343 financial and utility firms	1964-1996	Dividends, market-to-book ratio, ratio of dividends to book value of common equity	Neutral

(continued)



Table 13 – Continued

Determinant	Author(s)	Type	Sample	Period	Control variables	Relationship
	Hart and Morgan (1977)	E	113 industries in the UK	1968	employment concentration ratio, change in money sales, ratio of advertising to sales, firm size, employment, ratio of imports to domestically produced sales, number of firms in the industry	Positive
	Hanel and St-Pierre (2002)	E	302 US American, Canadian, and Japanese firms	1978-1991	Market share, price elasticity of demand, R&D capital intensity, various variables characterizing the firm's technological environment	Positive
Age	Coad et al. (2013)	E	Unbalanced panel of 73,891 Spanish manufacturing firms	1998-2006	Firm size, short-term and long-term debt ratios	Negative
	Majumdar (1997)	E	1,020 Indian firms	1988-1994	Firm size, liquidity, capital intensity, firm growth, debt-to-equity ratio, import ratio, diversity index	Negative
	Rosenbusch et al. (2011)	M				Negative

(continued)

Table 13 – Continued

Determinant	Author(s)	Type	Sample	Period	Control variables	Relationship
	Ito and Fukao (2010)	E	Approximately 14,000 Japanese overseas affiliates	1989-2002	Firm size, local sales ratio, Japanese equity ratio, local procurement ratio, WTO membership, management structure, place of business	Positive
	Qian et al. (2008)	E	Largest 189 US Fortune 500 firms	1996-2000	Firm size, leverage, regional diversification index, multinationality index, product scope index	Neutral
Productivity	Jovanovic (1982)	T				Positive: firms, which have incomplete information about their true costs and productivity levels, learn about their situation when operating in the market. A selection process leads to growth and survival of more productive and profitable firms, and to decline and exit of less efficient and profitable entities
	Demsetz (1973)	E	95 US industries	1963		Positive

(continued)

Table 13 – Continued

Determinant	Author(s)	Type	Sample	Period	Control variables	Relationship
	Bottazzi et al. (2008)	E	Unbalanced panel of 14,000-17,000 Italian manufacturing and 10,000-13,000 service firms	1998-2003	Credit ratings, sectoral affiliation	Positive
	Foster (1977)	T				Positive: the profitability threshold above which firms decide to enter the market depends positively on productivity

*Note:* Abbreviations in the third column refer to the type of the study. I distinguish E: empirical, T: theoretical, and M: meta-analysis.

## C.2 ESTIMATION

### C.2.1 Likelihood ratio test

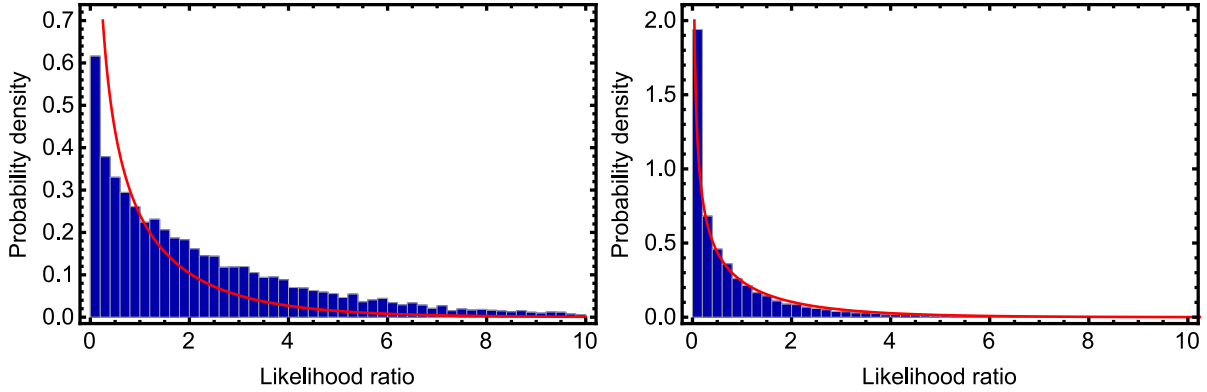


Figure 37: Simulated sampling distribution of the likelihood ratio test statistic (histogram) for the location (left) and dispersion parameter (right), superimposed with a  $\chi^2$  distribution with one degree of freedom (solid line). We observe that the  $\chi^2$  distribution is a reasonable description for the dispersion parameter, while it turns out to be too restrictive for the location parameter. This motivates why I use a data-dependent distribution of the test statistic for  $m$ .

### C.2.2 Numerical analyses on $m$ and $\sigma$ estimators

In this section, I attempt to quantify bias and estimation error in the maximum likelihood estimation of the Fokker-Planck equation. In particular, I focus on the effect of small samples, the adjustment speed of the diffusion, and the approximation error arising from the availability of discretely observed data. I simulate 500 realizations of the process

$$X_{t+\delta t} = X_t - \frac{D}{2\sigma} \text{sign}(X_t - m) \delta t + \sqrt{D\delta t} \cdot \eta_{t+\delta t}, \quad (34)$$

with different time series length  $T$  and location and scale parameters equal to their phenomenological values using the Euler-Maruyama method. In equation (34),  $\eta \sim \mathcal{N}(0,1)$  are independent and identically distributed standard normal random numbers, and  $\delta t = \Delta t/S$  are time increments with  $\Delta t = 1$  and micro time steps  $S \in \mathbb{N}^+$ . That is, the discrete time interval  $\Delta t$  is partitioned into  $S$  equal subintervals of width  $1/S$ . Then I re-estimate the location and dispersion parameter for each realization and compute three summary statistics. Let  $\theta$  denote the parameter to estimate and  $\theta_0$  is the correct value of this parameter. Then I consider the sample mean

$$\bar{\theta} = E[\hat{\theta}], \quad (35)$$

the standard error

$$s_{\hat{\theta}} = \sqrt{E[(\hat{\theta} - \bar{\theta})^2]}, \quad (36)$$

and the root mean squared error

$$\theta_{\text{RMSE}} = \sqrt{E[(\hat{\theta} - \theta_0)^2]} = \sqrt{E[(\hat{\theta} - \bar{\theta})^2] + (\bar{\theta} - \theta_0)^2}, \quad (37)$$

which is determined by the variance of the estimate and its bias. Table 14 summarizes the simulation results. It turns out that the estimate of the location parameter is unbiased also for small samples, irrespective of the characteristic time scale of the diffusion. The estimate of the dispersion parameter, on the other hand, exhibits a positive small sample bias that is most pressing when the adjustment speed is high ( $D = 10^{-2}$ ). Moreover, for small values of the diffusion coefficient ( $D = 10^{-4}$ ) and short time series ( $T = 30$ ), the standard error associated to both estimators is rather high and declines only slowly with sample size, which also explains the high dispersion of the estimates in Figure 31.

Table 14: Parameter estimates of simulated data as a function of the diffusion coefficient (D), time series length (T), and the number of micro time steps (S).

D	S	T	m			σ		
			Mean ( $\bar{m}$ )	SE ( $s_{\bar{m}}$ )	RMSE ( $m_{RMSE}$ )	Mean ( $\bar{\sigma}$ )	SE ( $s_{\bar{\sigma}}$ )	RMSE ( $\sigma_{RMSE}$ )
0.0001	1	30	0.097	0.079	0.079	0.064	0.051	0.051
		100	0.099	0.072	0.072	0.070	0.049	0.050
		200	0.097	0.071	0.071	0.070	0.047	0.048
		500	0.097	0.061	0.061	0.070	0.039	0.041
		1000	0.093	0.037	0.037	0.067	0.028	0.029
	10	30	0.093	0.075	0.075	0.066	0.049	0.050
		100	0.092	0.073	0.073	0.068	0.047	0.048
		200	0.093	0.065	0.065	0.067	0.046	0.047
		500	0.092	0.051	0.051	0.066	0.035	0.036
		1000	0.094	0.038	0.038	0.065	0.032	0.032
	100	30	0.097	0.077	0.077	0.064	0.051	0.052
		100	0.090	0.076	0.076	0.067	0.049	0.050
		200	0.100	0.063	0.063	0.068	0.043	0.044
		500	0.095	0.059	0.059	0.071	0.043	0.045
		1000	0.094	0.044	0.044	0.064	0.031	0.032
0.001	1	30	0.094	0.071	0.071	0.075	0.046	0.049
		100	0.098	0.044	0.044	0.075	0.031	0.036
		200	0.096	0.030	0.030	0.072	0.022	0.026
		500	0.095	0.012	0.012	0.071	0.012	0.018
		1000	0.096	0.008	0.008	0.071	0.010	0.016
	10	30	0.097	0.070	0.070	0.070	0.046	0.048
		100	0.094	0.042	0.042	0.068	0.031	0.032
		200	0.096	0.023	0.023	0.065	0.021	0.022
		500	0.095	0.009	0.009	0.061	0.011	0.012
		1000	0.095	0.005	0.005	0.060	0.007	0.007
	100	30	0.094	0.071	0.071	0.069	0.045	0.046
		100	0.094	0.031	0.031	0.063	0.026	0.026
		200	0.096	0.022	0.022	0.062	0.018	0.019
		500	0.095	0.007	0.007	0.061	0.012	0.012
		1000	0.095	0.004	0.004	0.059	0.008	0.008
0.01	1	30	0.096	0.046	0.046	0.110	0.028	0.059
		100	0.095	0.021	0.021	0.113	0.015	0.057
		200	0.096	0.014	0.014	0.113	0.012	0.056
		500	0.095	0.007	0.007	0.113	0.007	0.055
		1000	0.095	0.005	0.005	0.113	0.005	0.055
	10	30	0.095	0.020	0.020	0.066	0.019	0.021
		100	0.094	0.011	0.011	0.065	0.010	0.012
		200	0.095	0.007	0.007	0.065	0.007	0.009
		500	0.095	0.005	0.005	0.064	0.004	0.008
		1000	0.095	0.004	0.004	0.065	0.003	0.007
	100	30	0.096	0.018	0.018	0.062	0.018	0.018
		100	0.095	0.008	0.008	0.060	0.009	0.010
		200	0.095	0.005	0.005	0.059	0.006	0.006
		500	0.095	0.003	0.003	0.059	0.004	0.004
		1000	0.095	0.003	0.003	0.059	0.003	0.003

Note: Results refer to 500 simulated realizations of the diffusion process with  $m = 0.095$  and  $\sigma = 0.058$ .

### C.3 ROBUSTNESS CHECKS

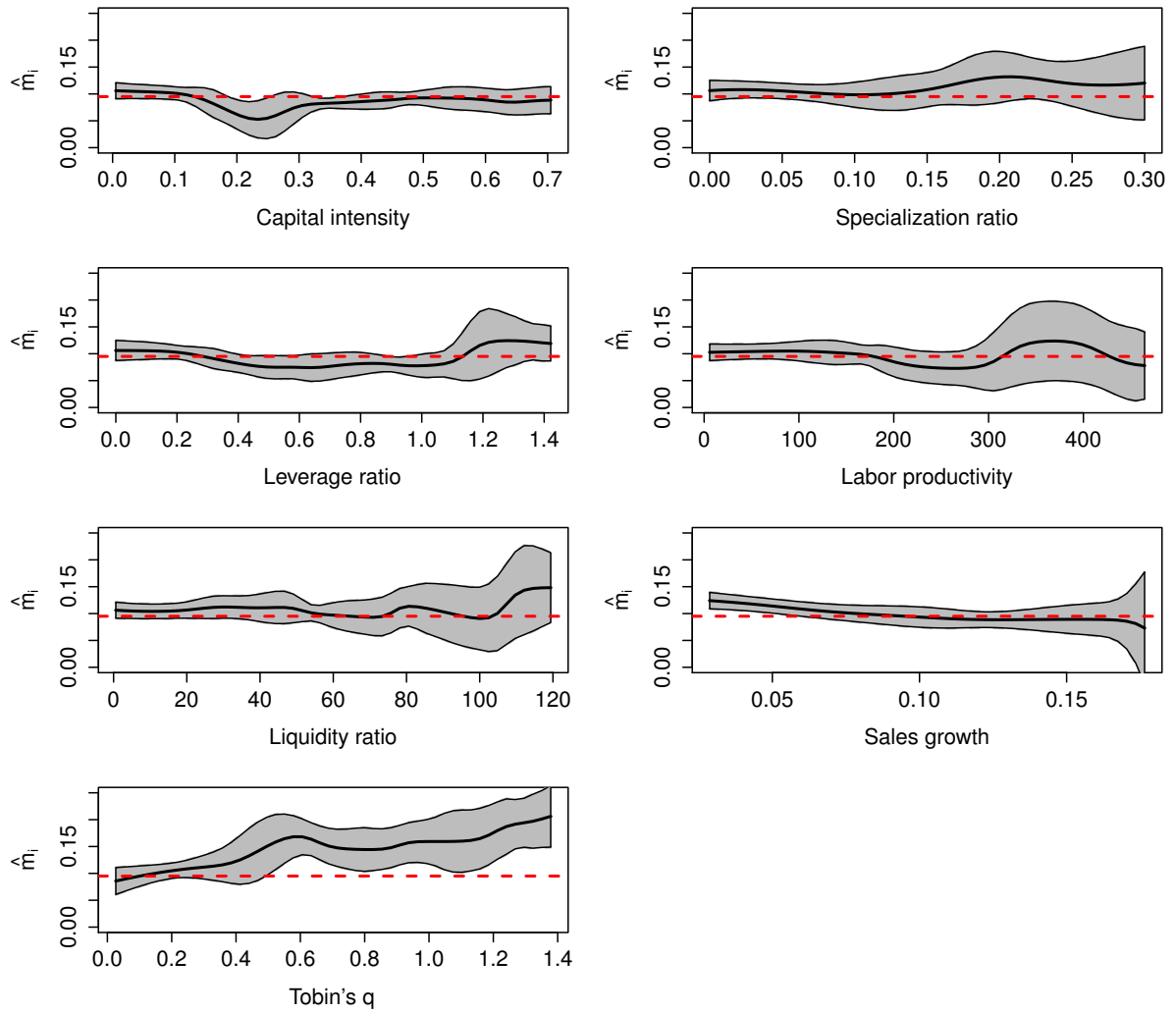


Figure 38: Relation between the location parameter and time series volatility of corporate characteristics, measured in terms of median absolute deviation. As before, dashed horizontal lines illustrate the phenomenological value  $\hat{m} = 0.0947$ .

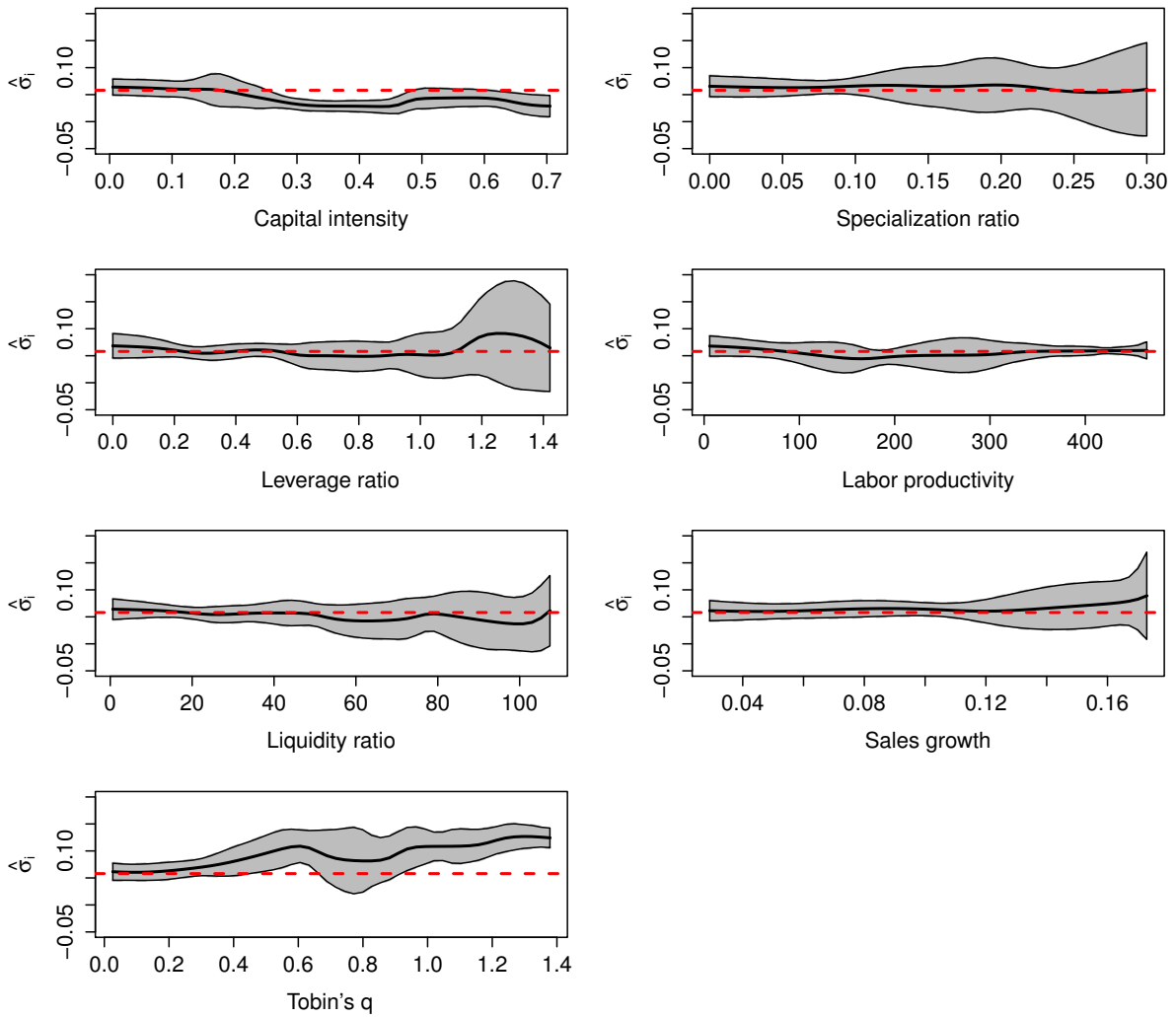


Figure 39: Relation between the dispersion parameter and time series volatility of corporate characteristics, measured in terms of median absolute deviation. As before, dashed horizontal lines illustrate the phenomenological value  $\hat{\sigma} = 0.0581$ .



#### C.4 EFFECT OF LIFESPAN ON AVERAGE PROFITABILITY AND ITS VOLATILITY

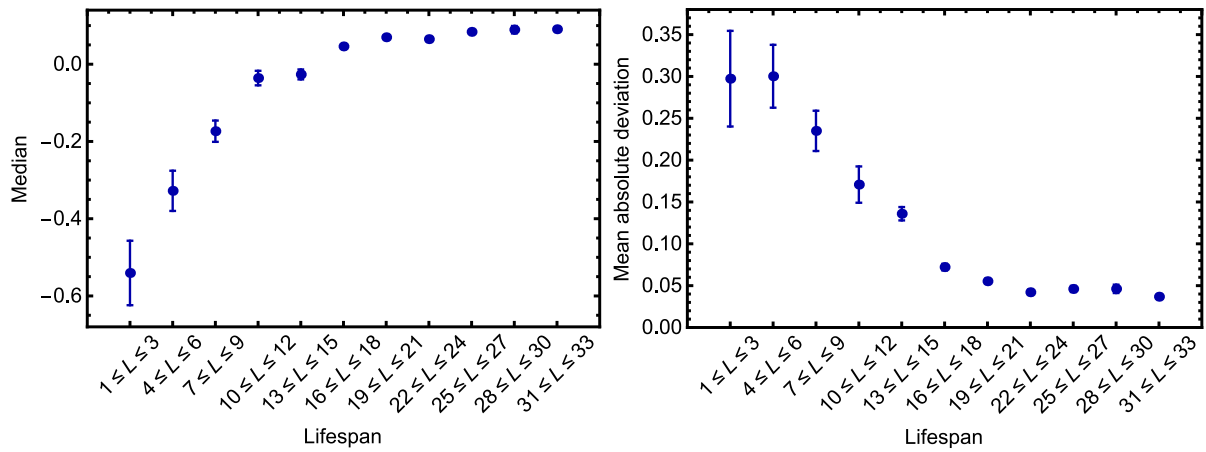


Figure 40: Average profitability (left) and its volatility (right) as a function of lifespan  $L$ . Lifespan is measured in terms of the number of years that firms survive at least in the sample. The points illustrate the median  $m_i$  and  $\sigma_i$ , respectively, of all firms in each bin. Error bars correspond to one standard error that has been computed as the standard deviation of 1000 bootstrap samples of the statistics in each bin.



## A STATISTICAL EQUILIBRIUM APPROACH TO FORECASTING CORPORATE PROFITABILITY

---

## ABSTRACT

This paper proposes a statistical equilibrium methodology to model and forecast firm profitability that accounts for the fat-tailed distribution of return on assets (ROA). The approach builds on a new type of data generating process (AMIK diffusion) that has recently been introduced by Alfarano et al. (2012). Employing the maximum likelihood method for the estimation of this process, I assess its out-of-sample forecasting performance at short and longer time horizons. The ability of the model to predict ROA is evaluated relative to another prominent mean-reverting diffusion process, namely the Ornstein-Uhlenbeck model, and other univariate time-series models from the ARMA and ARIMA varieties. Using balance sheet data of more than 500 long-lived publicly traded US corporates, I find that the AMIK diffusion is not outperformed by other time series models in most cases if we add the assumption that profit rates are ergodic.

## 4.1 INTRODUCTION

Forecasts of profitability are relevant for firm managers, financial analysts, investors, and other groups of corporate stakeholders. Amongst others, they may affect corporate strategy, investment decisions, and the valuation of equity (see, e.g., the residual income model by Feltham and Ohlson, 1995; Ohlson, 1995). The perception that, to some extent, profitability is predictable arises because accounting rates of return, unlike financial returns, are persistent and exhibit a mean-reverting property in their time series (Fama and French, 2000; Freeman et al., 1982; Nissim and Penman, 2001). From a methodological viewpoint, this raises the question of how to model the dynamics of profitability most appropriately.

Despite the appealing statistical properties of profit rates, it appears that profitability has received relatively little attention in the forecasting literature, mostly in accounting and finance, and that the majority of existing papers is concerned with modeling and forecasting of earnings as an alternative measure of firm performance. Several early contributions from the 1960/70s explore the time-series properties of earnings. They come to the conclusion that earnings follow a martingale process, implying that the best prediction for earnings is the “naive” forecast (e.g. Ball and Watts, 1972; Lintner and Glauber, 1978; Little, 1962). Other studies investigate the forecasting capacity of various autoregressive (integrated) moving average models in the spirit of Box and Jenkins (1970) (e.g. Albrecht et al., 1977; Brown and Rozeff, 1979; Callen et al., 1993; Collins and Hopwood, 1980; Foster, 1977; Griffin, 1977; Lookabill, 1976; Watts and Leftwich, 1977). In the research field of corporate profitability, the majority of studies employ, instead of time-series models, regression analysis that postulates some sort of structural relationship between profitability and other variables. For example, Fairfield et al. (1996) analyze the predictive content of several earnings components (e.g. operating earnings, non-operating earnings and taxes, and special items) on the return on equity and find that disaggregation improves the forecasting

accuracy relative to the models subject to the higher level of aggregation. In a similar vein, Fairfield and Yohn (2001), Soliman (2008), and Bauman (2014) use the DuPont methodology to decompose ROA into the product of asset turnover and profit margin. They provide evidence that changes in asset turnover and profit margin contain valuable information for forecasting the change in ROA. Employing a variant of the Fama and MacBeth (1973) regression, Fama and French (2000) and Allen and Salim (2005) conduct forecasting analysis on ROA in order to draw insights into the change of earnings. Their model uses the ratio of dividends to the book value of equity, the market-to-book ratio, and a dummy variable to distinguish between dividend and non-dividend paying firms as explanatory variables for expected profitability. One of the few exceptions that employs time-series analysis to forecast profitability is the study by Fairfield et al. (2009) which attempts to predict returns on equity and net operating assets by means of autoregressive processes. Moreover, although time series models have been seldomly used to forecast profitability, they play the dominant role in the rich persistence of profit literature that employs autoregressive processes as their major work horse (e.g. Geroski and Jacquemin, 1988; Goddard and Wilson, 1999; Gschwandtner, 2005; Mueller, 1986, 1990).

In this paper, I employ the statistical equilibrium methodology that has recently been introduced by Alfarano and Milaković (2008) and Alfarano et al. (2012), and extend the scope of its relevance by examining the forecasting power of this approach on ROA. The model presumes that profitability is subject to ergodicity conditional on survival, i.e. that time-series and cross-sectional properties of the data will converge on a sufficiently long time scale. This presumption makes it feasible to extract information on the individual time-series from the cross-sectional profit rate distribution and leads to forecasting gains when more data is available in the cross-sectional sphere than in the time-series dimension, which is frequently the case in practical applications. The perception that the ensemble average may disclose information on the individual time-series has been prevalent also in previous investigations. For instance, Fairfield et al. (2009) report the superiority of models predicting mean-reversion to economy-wide rather than industry-specific levels. Yet, by its very nature, their autoregressive model cannot account for several features of the data. Instead, the statistical equilibrium approach offers, at least, three reasons why the model based on it should be a valuable tool to forecast profitability. First, it reproduces the empirical distribution of profit rates that is markedly non-Gaussian but approximately follows the Laplace distribution (see e.g. Alfarano and Milaković, 2008; Alfarano et al., 2013; Scharfenaker and Semieniuk, 2016), implying that small deviations around the mean and extreme events occur more often than in the case of a normal distribution. Second, the model is also consistent with the autocorrelation structure of the data that exhibit an (asymptotic) exponential decay, i.e. short memory, as pointed out by Mundt et al. (2016). Finally, Mundt (2017) presents evidence that individual firm characteristics are almost negligible for the dynamics of profitability if the entity has survived in the market for an extended period of time, which testifies to the existence of a common law of motion governing profitability for long-lived firms. Therefore, it should be worthwhile to investigate if the model's ability to replicate these features also materializes in superior

forecasting capacity. To this end, I estimate the stochastic process of firm profitability via maximum likelihood and a closed-form solution to its transient density using a sample of more than 500 publicly traded US non-banking corporations, and assess the model's goodness of fit with respect to its out-of-sample forecasting performance at short and longer time horizons (up to three years ahead). The performance of the model is evaluated relative to another prominent mean-reverting stochastic process, namely the Ornstein-Uhlenbeck model, and more established time-series models from the mixed autoregressive integrated moving average (ARIMA) family. Since the significance of its properties has been widely discussed in the literature on forecasting earnings, I also consider the random walk process as a special case of an ARMA model. In the paper, comparison of forecasting power between all competing models hinges on two alternative loss functions, mean squared error (MSE) and mean absolute error (MAE), as well as the superior predictive ability (SPA) test by Hansen (2005).

While it is, of course, statistically convenient to select firms with a long history of data in order to maximize the number of observations, the main reason for choosing this particular sample goes far beyond technical considerations. In fact, the selection of long-lived firms is inevitable because the notion of statistical equilibrium prevails merely for surviving entities. Shorter lived firms, i.e. entities that have only recently entered the market or disappear within the sample period, are not subject to the common mechanism that implies a kind of auto-pilot mode for those which survive. Why should it be relevant to study these corporations? Interestingly, it turns out that surviving firms are often large in terms of size: the sample used for the present investigation contains more than 200 entities that are listed on the 2013 Forbes Fortune 500 list. Given that gross revenues of the largest 500 US corporations accounted for approximately 73 percent of US nominal GDP in 2013, their impact on the overall economy is certainly non-negligible and their individual destinies may have even macroeconomic implications (Gabaix, 2011, provides empirical evidence on this "granular hypothesis"). Thus, surviving corporations are a very relevant group of firms to study.

One of the key results obtained in this paper is that the statistical equilibrium model is not outperformed by other models in most cases, under both the MSE and MAE loss function. For one year ahead forecasts, the random walk is the toughest competitor, while the Ornstein-Uhlenbeck process turns out to be second-best for longer horizons. Box-Jenkins ARMA and ARIMA-type models are clearly dominated by the former two diffusion processes. Further, perhaps somewhat surprisingly, the random walk outperforms models predicting mean-reversion to firm-specific profitability levels. While this respectable overall performance of the naïve forecast illustrates the difficulty of predicting individual behavior in a system of interacting individuals, it also implies that the notion of statistical equilibrium is essential to realize forecasting gains.

The remainder of this paper is organized as follows. Section 4.2 describes the data and outlines the forecasting design. Section 4.3 presents the statistical equilibrium

Table 15: Sample composition by sectors.

Division	SIC codes	Number of firms
Agriculture, forestry, and fishing	01-09	1
Mining	10-14	21
Construction	15-17	9
Manufacturing	20-39	279
Transportation and public utilities	40-49	87
Wholesale trade	50-51	16
Retail trade	52-59	33
Finance, insurance, and real estate	62-67	32
Services	70-89	38
All		516

*Note:* Firms are classified according to the business segments that provided the highest revenue at the end of 2013.

model and introduces its competitors. Section 4.4 reports the main results and section 4.5 summarizes and concludes.

#### 4.2 DATA AND FORECASTING METHODOLOGY

The dataset for this study is obtained from Thomson Reuters and builds on firms' annual financial statements, as reported in the Datastream Worldscope Database. Worldscope includes publicly quoted companies and provides the best coverage and longest history of data for developed markets in North America, where the earliest information is available for 1980. Since this investigation aims at long-lived firms, my focus will be on US companies that are listed on the stock exchange since 1980 and still exist in 2013, the year when the data for this study were collected. This condition is met by 516 entities operating in virtually all sectors of the US economy except banking, as shown in Table 15.

Profitability of these firms is measured in terms of return on assets, computed as the ratio of the flow of operating income to the stock of total assets, which serves as an approximation to the profit rate. Table 16 provides several summary statistics for this financial ratio. They suggest that the cross-sectional profit rate distribution is fairly symmetric around the mean. The year-by-year skewness statistics do not exhibit any clear pattern, indicating that neither negative nor positive skew is an universal feature of the data. Negative realizations of the skewness statistic occurred mainly during the last financial and banking crisis, yet it turns out that this is due to extremely few observations. Moreover, the annual cross-sectional profit rate distributions are markedly non-Gaussian since they exhibit considerable excess kurtosis, i.e. fatter tails than the Normal distribution. This is confirmed by the Anscombe and Glynn (1983) test which clearly rejects the null hypothesis of zero excess kurtosis at any level of significance (see Table 17). Various goodness-of-fit tests, also summarized in Table 17, support the perception that the data are not normally distributed. They reject the null

Table 16: Summary statistics for the cross-sectional profit rate distribution.

Year	Mean	SD	Skewness	Kurtosis	Min	Max	Obs.
1980	0.1408	0.1039	2.4947	21.0385	-0.2581	0.9585	516
1981	0.1345	0.0939	3.3124	36.7539	-0.1765	1.2094	516
1982	0.1172	0.0876	0.8289	6.3440	-0.1421	0.5749	516
1983	0.1188	0.0893	1.2978	10.8612	-0.1915	0.7362	516
1984	0.1267	0.0795	0.4899	6.6484	-0.2442	0.5687	516
1985	0.1148	0.0809	0.8484	7.5380	-0.1450	0.6072	516
1986	0.1025	0.0887	0.1384	7.8971	-0.2771	0.6466	516
1987	0.1044	0.0841	-0.8784	14.6086	-0.5934	0.5220	516
1988	0.1066	0.0789	-0.6874	10.0193	-0.4856	0.3465	516
1989	0.1064	0.0732	0.0756	4.7722	-0.1739	0.3855	516
1990	0.1001	0.0721	-0.1188	6.3441	-0.3135	0.3931	516
1991	0.0905	0.0721	-0.1145	6.6357	-0.3169	0.4282	516
1992	0.0930	0.0697	0.4469	5.0058	-0.1655	0.4052	516
1993	0.0957	0.0668	0.6979	5.6224	-0.1413	0.4220	516
1994	0.1013	0.0740	-0.3734	12.1097	-0.3714	0.4906	516
1995	0.1048	0.0734	-0.6240	11.0761	-0.4347	0.4531	516
1996	0.1057	0.0696	0.0404	6.8725	-0.2435	0.4478	516
1997	0.1071	0.0732	0.1539	9.1781	-0.2950	0.5896	516
1998	0.1050	0.0692	0.0613	7.5899	-0.2895	0.4689	516
1999	0.1011	0.0716	0.0740	5.3818	-0.1854	0.3948	516
2000	0.0999	0.0812	0.8255	12.6345	-0.2969	0.7243	516
2001	0.0849	0.0850	1.3233	20.7017	-0.4831	0.8170	516
2002	0.0803	0.0862	0.3010	29.3674	-0.6976	0.8626	516
2003	0.0830	0.0815	0.3818	21.9214	-0.6107	0.7044	516
2004	0.0924	0.0803	1.8493	20.1981	-0.2525	0.8285	516
2005	0.0944	0.0841	-0.5844	15.4016	-0.5064	0.6448	516
2006	0.0996	0.0837	-0.0314	19.4832	-0.5756	0.7337	516
2007	0.0932	0.0918	-1.4965	22.2446	-0.7486	0.6888	516
2008	0.0839	0.1389	-9.5543	159.0419	-2.2566	0.5060	516
2009	0.0685	0.0920	-1.9882	21.7573	-0.8252	0.3883	516
2010	0.0854	0.0887	-2.6070	42.1125	-0.9420	0.6562	516
2011	0.0914	0.1018	-2.3489	79.0333	-1.2427	1.1218	516
2012	0.0861	0.1110	1.0702	57.0922	-0.9368	1.3831	516
2013	0.0865	0.1284	-0.3022	130.0858	-1.6702	1.7950	516



Table 17: P-values of normality and kurtosis tests.

Year	AD	CVM	KUI	JB ALM	SW	AG
1980	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1981	0.0000	0.0004	0.0000	0.0000	0.0000	0.0000
1982	0.0113	0.0200	0.0010	0.0000	0.0000	0.0000
1983	0.0022	0.0084	0.0000	0.0000	0.0000	0.0000
1984	<b>0.0507</b>	<b>0.0608</b>	0.0042	0.0000	0.0000	0.0000
1985	0.0108	0.0219	0.0003	0.0000	0.0000	0.0000
1986	0.0004	0.0011	0.0000	0.0000	0.0000	0.0000
1987	0.0001	0.0002	0.0000	0.0000	0.0000	0.0000
1988	0.0015	0.0033	0.0000	0.0000	0.0000	0.0000
1989	0.0107	0.0126	0.0000	0.0000	0.0000	0.0000
1990	0.0028	0.0037	0.0000	0.0000	0.0000	0.0000
1991	0.0093	0.0130	0.0000	0.0000	0.0000	0.0000
1992	0.0144	0.0183	0.0000	0.0000	0.0000	0.0000
1993	0.0016	0.0021	0.0000	0.0000	0.0000	0.0000
1994	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
1995	0.0011	0.0032	0.0000	0.0000	0.0000	0.0000
1996	0.0048	0.0100	0.0000	0.0000	0.0000	0.0000
1997	0.0030	0.0085	0.0000	0.0000	0.0000	0.0000
1998	0.0061	0.0150	0.0001	0.0000	0.0000	0.0000
1999	0.0058	0.0153	0.0000	0.0000	0.0000	0.0000
2000	0.0004	0.0022	0.0000	0.0000	0.0000	0.0000
2001	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
2002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2005	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
2006	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2007	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2009	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2011	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2012	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2013	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

*Note:* Abbreviations refer to AD: Anderson-Darling test (Anderson and Darling, 1952), CVM: Cramér-Von Mises test (Cramér, 1928; Mises, 1928), KUI: Kuiper test (Kuiper, 1960), JB ALM: Jarque-Bera adjusted Lagrange multiplier test (Urzúa, 1996), Shapiro-Wilk test (Shapiro and Wilk, 1965), and Anscombe-Glynn test (Anscombe and Glynn, 1983). P-values greater than 5 percent are shown in boldface. Entries equal to 0.0000 imply p-value  $< 5 \times 10^{-5}$ .

of Gaussian distributed profit rates for all except one of the 34 annual cross-sectional distributions on the 5 percent level. Altogether, these empirical observations support the hypothesis of a symmetric leptokurtic profit rate distribution as embodied in the AMIK model.

Turning to the time-series properties, the autocorrelation functions in Figure 45 in Appendix D.1 suggest that profit rates are highly serially correlated. This graphical impression is confirmed by the Ljung and Box (1978) and Box and Pierce (1970) tests which both reject the null hypothesis of zero autocorrelations in approximately 450 out of 516 cases on the 5 percent level. Accordingly, profit rates do not move erratically but appear predictable, to some extent, due to the rich statistical structure and memory that we find in the data. Finally, I have also tested for stationarity of individual profit rate time-series. Clearly, such formal testing is hampered by the rather small number of observations available in the time domain and the fact that profit rates are positively autocorrelated, in particular if the diffusion of profit rates is rather slow. Yet, on the 5 percent level, I cannot reject the null hypothesis of (second order) stationarity in approximately 70 percent of cases based on the Priestley and Rao (1969) test which considers time-variations in the Fourier spectrum. It is worth noting that a similar frequency is observed for synthetic data of the same size that obey the AMIK diffusion, and that the test easily detects stationarity in simulated data as the length of the time-series is growing larger. I take this to imply that the remaining, approximately 150, profit rate time series are not necessarily non-stationary, but rather attribute this result to the limited power of the test for small samples.<sup>44</sup> Moreover, notice that both mean and standard deviation of the cross-sectional profit rate distribution exhibit remarkably little fluctuations over time, which is already some indication for stationarity of the data.

#### 4.2.1 *Forecast design*

To forecast profitability, I employ a rolling window scheme where I estimate a model in-sample and use the fitted model to obtain out-of-sample predictions for forecast horizons up to 3 years ahead. The rolling window permits to generate several predictions for each forecast horizon from a single time-series, and is less sensitive to effects in single years than the fixed scheme because forecast errors are averaged across different time periods. Moreover, the SPA test does not allow for parameters that are estimated with a recursive scheme, which is why I do not consider it here.<sup>45</sup> In order to obtain and evaluate the forecasts, I split each time-series into two subsamples. The first one consists of 20 annual observations and serves as in-sample or training period for parameter estimation, while the following 3 observations are used as out-of-

---

<sup>44</sup> This problem is also relevant for other popular tests for (non-)stationarity, e.g. those which test for unit roots in autoregressions (see, e.g., the discussion in Cochrane, 1991).

<sup>45</sup> While the fixed scheme considers an in-sample of fixed size, the recursive scheme employs a window of in-sample observations that includes all data points from the beginning of the in-sample period and successively adds more recent observations. Thus, the recursive scheme implies an in-sample that is expanding over time.

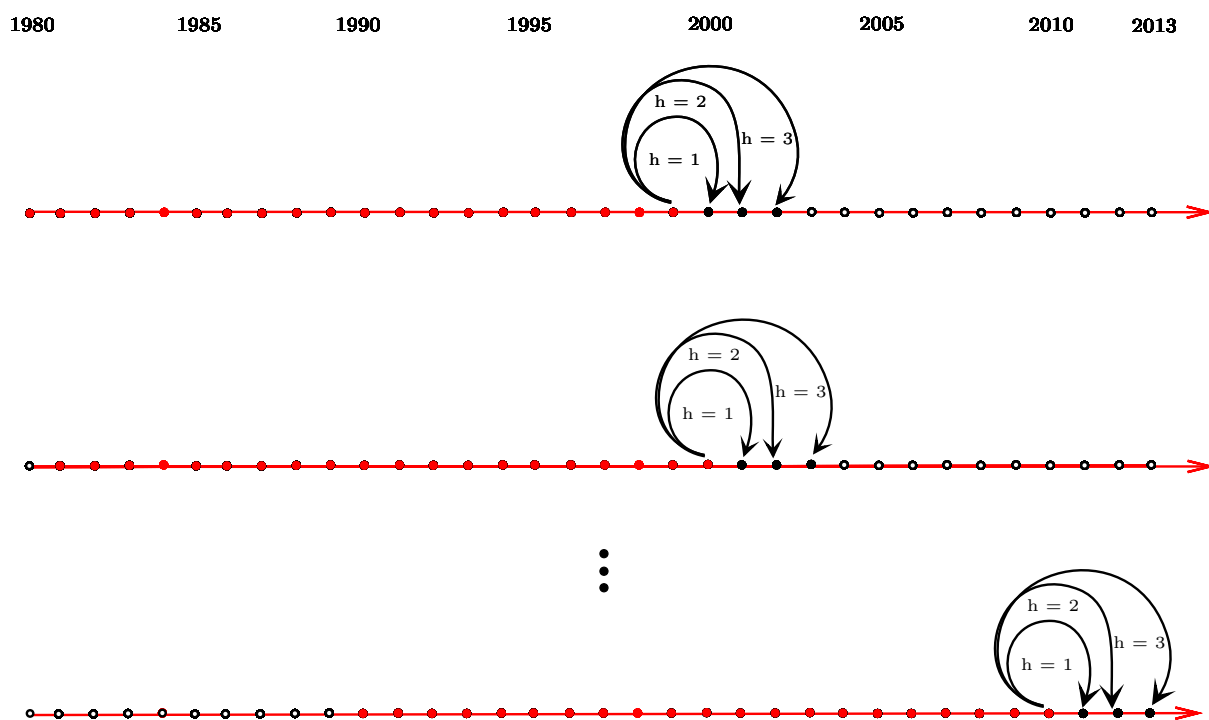


Figure 41:  $h$  step ahead forecasts with a rolling window.

sample or testing period for forecast evaluation. Both the training and testing sample are then rolled forward by adding one more recent observation and dropping one from the beginning of the respective period, so that the number of observations in each of the two subsamples remains constant. Given that no data is available after 2013, this scheme implies that the rolling window can be shifted 11 more times, permitting a total of 12 predictions for each forecast horizon  $h = 1, 2, 3$  years. Figure 41 provides an illustration of the rolling window scheme.

#### 4.2.2 Forecast evaluation

The relative out-of-sample predictive accuracy of each model is evaluated by means of the test outlined in Hansen, 2005. Unlike the prominent Diebold and Mariano, 1995 test that examines if two models exhibit equal predictive accuracy, Hansen's approach provides a testing framework for superior predictive ability (SPA) that compares the forecasting error of a single (benchmark) model relative to the whole set of competitors. It constitutes a refinement of the reality check (RC) for data snooping that has been suggested by White, 2000, considering a studentized test statistic and a sample-dependent null distribution as modifications. The bootstrap should control for the small sample at our disposal, and avoids reference to asymptotic results which are potentially misleading in our case due to the relatively small number of observations. In what follows, I give a brief summary of the SPA test.

Hansen's test for SPA asks whether any of the alternatives is superior to the benchmark model in terms of forecast accuracy. To this end, it considers the null hypothesis that the benchmark is not outperformed by any of the competitors in terms of expected loss. To fix notation, let  $x_{t+h}$  denote the ex-post realized return on assets in period  $t + h$ , and  $\hat{x}_{t+h}$  is the corresponding forecast. Then the expression

$$d_{k,t+h} \equiv L(x_{t+h}, \hat{x}_{0,t+h}) - L(x_{t+h}, \hat{x}_{k,t+h}) \quad (38)$$

quantifies the loss of competitor  $k = 1, \dots, m$  at time  $t + h$  relative to the benchmark model  $0$ , where  $L$  denotes some real-valued loss function, e.g. squared or absolute forecasting errors. Relative losses of all competitors are collected in the vector

$$\mathbf{d}_{t+h} = (d_{1,t+h}, \dots, d_{m,t+h})'. \quad (39)$$

If the benchmark is not outperformed by other models, the expected relative loss of each competitor,  $\mu_k \equiv E[d_{k,t+h}]$ , is smaller than or equal to zero.<sup>46</sup> Therefore, the null hypothesis is given by

$$H_0 : \boldsymbol{\mu} \leq \mathbf{0}, \quad (40)$$

where  $\boldsymbol{\mu} \equiv (\mu_1, \mu_2, \dots, \mu_m)'$ . This hypothesis is then tested via the test statistic

$$T^{SPA} \equiv \max_{k=1, \dots, m} \left[ \frac{\sqrt{N}\bar{d}_1}{\hat{\omega}_1}, \dots, \frac{\sqrt{N}\bar{d}_m}{\hat{\omega}_m}, 0 \right], \quad (41)$$

where  $N$  denotes the number of observations that are employed for model comparison,

$$\bar{d}_k = N^{-1} \sum_{t=1}^N d_{k,t+h} \quad (42)$$

is the *sample* mean loss differential of model  $k$ ,  $\omega_k^2 \equiv \text{Var}(\sqrt{N}\bar{d}_k)$ , and  $\hat{\omega}_k^2$  denotes an estimate of  $\omega_k^2$  (see Hansen, 2005, p. 372 for details on the estimation). A major difference from White's RC is that the test statistic in equation (41) employs studentized loss differentials which is supposed to improve the power of the test if forecasting performances exhibit cross-sectional heteroskedasticity.<sup>47</sup>

Finally, the distribution of the test statistic under the null hypothesis is approximated by resampling the vector of relative performances,  $\mathbf{d}_{t+h}$  in equation (39), using

<sup>46</sup> Notice that  $\mu_k$  refers to the (unknown) *population* mean of the loss differential series.

<sup>47</sup> Dividing estimates of average loss differentials by their variability (studentization) ensures that losses of all models are compared in the same units of standard deviation. Thus, models with very volatile performances compare less favorably (see Hansen, 2005, p. 369).

the stationary bootstrap proposed by Politis and Romano (1994). These bootstrap series are denoted by

$$\mathbf{d}_{b,t+h}^* = (d_{1,b,t+h}^*, \dots, d_{m,b,t+h}^*)', \quad (43)$$

where  $b = 1, \dots, B$ , and  $B$  is the number of bootstrap samples of length  $N$ . Then, the null is imposed by recentering the performance differentials in  $\mu$ . Since poor models could potentially bias the test result towards acceptance of the null hypothesis, Hansen (2005) suggests to exclude irrelevant alternatives when simulating the null distribution. Here, poor means that the studentized average loss differential,  $n^{1/2}\bar{d}_k/\hat{\omega}_k$ , is smaller than the threshold  $\sqrt{2\ln(\ln(N))}$ . Formally, this is achieved by defining the centered variable

$$Z_{k,b,t+h}^* \equiv d_{k,b,t+h}^* - \bar{d}_k \cdot 1_{\{n^{1/2}\bar{d}_k/\hat{\omega}_k \geq -\sqrt{2\ln(\ln(N))}\}}, \quad (44)$$

where  $1$  denotes the indicator function. The average of the left-hand side of equation (44) is

$$\bar{Z}_{k,b}^* = N^{-1} \sum_{t=1}^N Z_{k,b,t+h}^*. \quad (45)$$

Since  $\bar{Z}_{k,b}^*$  is the centered bootstrap analog of  $\bar{d}_k$ , the distribution of the test statistic under the null is derived from the bootstrap test statistic

$$T_b^{*SPA} \equiv \max_{k=1, \dots, m} \left[ \frac{\sqrt{N}\bar{Z}_{1,b}^*}{\hat{\omega}_1}, \dots, \frac{\sqrt{N}\bar{Z}_{m,b}^*}{\hat{\omega}_m}, 0 \right]. \quad (46)$$

Finally, the estimated p-value of the test statistic  $T^{SPA}$  is given by the percentage of bootstrap test statistics  $T_b^{*SPA}$  that exceed  $T^{SPA}$ , i.e.

$$\hat{p}^{SPA} \equiv B^{-1} \sum_{b=1}^B 1_{T_b^{*SPA} > T^{SPA}}. \quad (47)$$

### 4.3 COMPETING MODELS

The following section introduces the competing forecasting models. The starting point is the AMIK diffusion which has recently been introduced as a model for the time evolution of long-lived firms' profit rates. Among the competitors I consider the Ornstein-Uhlenbeck process as an alternative mean-reverting diffusion, as well as more established time-series models of the mixed autoregressive and moving average varieties. Details about model selection, estimation, and forecasting are provided in each subsection.

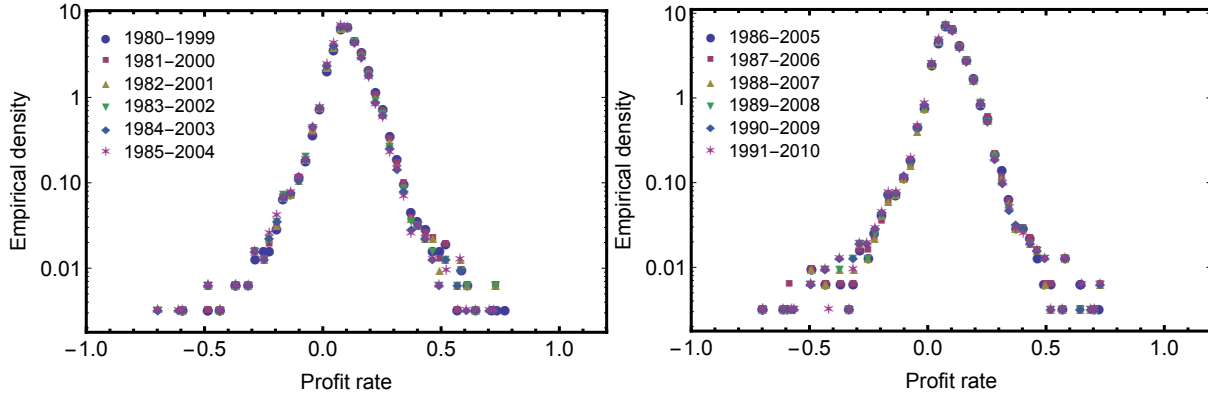


Figure 42: Pooled empirical density of annual profit rates for 516 long-lived publicly traded US companies vs. Laplace fit for different in-sample periods.

#### 4.3.1 Statistical equilibrium model of competitive firms

Alfarano et al. (2012) and Alfarano and Milaković (2008) have recently proposed a statistical equilibrium model of competitive firms that approaches firm competition from a probabilistic perspective. They argue that the tendency of competition to equalize returns to capital across different uses leads to an *equilibrium distribution* of profit rates that characterizes the macroscopic properties of the ensemble of interacting business firms. Given that this cross-sectional distribution is fairly symmetric but non-Gaussian due to excess kurtosis, they suggest to employ a more fat-tailed model and fit a Laplace distribution

$$f_S(x) = \frac{1}{2\sigma} \exp\left(-\left|\frac{x - m}{\sigma}\right|\right), \quad (48)$$

with location  $m \in \mathbb{R}$  and scale parameter  $\sigma \in \mathbb{R}^+$ . A battery of goodness-of-fit tests in Table 18 suggest that the Laplace distribution is indeed a reasonable, albeit not perfect, approximation of the data since the null hypothesis of Laplacian distributed profit rates cannot be rejected on the 5 percent level in the majority of cases.<sup>48</sup> Following this approach, I fit the cross-sectional profit rate distribution with equation (48). Figure 42 depicts the empirical density for the different portions of in-sample data, while the corresponding maximum likelihood (ML) estimates are provided in Table 19. Plots of the empirical densities corroborate the Laplace hypothesis since a log-linear plot shows the characteristic tent-shape.

<sup>48</sup> I very much suspect that some departure from the Laplace distribution can be explained with earnings management which leads to anomalies of profit rate realizations around  $x \approx 0$ . This is a well documented phenomenon in the accounting literature (Burgstahler and Dichev, 1997).

Table 18: P-values of various goodness-of-fit tests for the Laplace distribution.

Year	AD	CVM	KS	KUI	WU <sub>2</sub>
1980	0.0344	<b>0.1001</b>	<b>0.0809</b>	0.0020	0.0064
1981	0.0049	0.0199	0.0190	0.0157	0.0024
1982	<b>0.0944</b>	<b>0.1533</b>	<b>0.1112</b>	<b>0.0907</b>	<b>0.1257</b>
1983	<b>0.1252</b>	<b>0.1845</b>	<b>0.1551</b>	0.0073	0.0265
1984	<b>0.3737</b>	<b>0.4675</b>	<b>0.4168</b>	<b>0.3815</b>	<b>0.2326</b>
1985	<b>0.2452</b>	<b>0.3143</b>	<b>0.1697</b>	<b>0.0989</b>	<b>0.1615</b>
1986	<b>0.6617</b>	<b>0.6492</b>	<b>0.7440</b>	<b>0.9412</b>	<b>0.8879</b>
1987	<b>0.5748</b>	<b>0.5212</b>	<b>0.3833</b>	<b>0.2245</b>	<b>0.3242</b>
1988	<b>0.2099</b>	<b>0.3568</b>	<b>0.4789</b>	<b>0.4196</b>	<b>0.5583</b>
1989	<b>0.0669</b>	<b>0.1310</b>	<b>0.1917</b>	<b>0.4221</b>	<b>0.3156</b>
1990	<b>0.3666</b>	<b>0.4832</b>	<b>0.4073</b>	<b>0.5415</b>	<b>0.5940</b>
1991	<b>0.6249</b>	<b>0.6611</b>	<b>0.8287</b>	<b>0.5187</b>	<b>0.4731</b>
1992	<b>0.3180</b>	<b>0.4917</b>	<b>0.4223</b>	<b>0.3509</b>	<b>0.4383</b>
1993	<b>0.0651</b>	<b>0.1763</b>	<b>0.1753</b>	<b>0.4651</b>	<b>0.3973</b>
1994	0.0389	<b>0.1073</b>	<b>0.1681</b>	<b>0.2117</b>	<b>0.2526</b>
1995	<b>0.2115</b>	<b>0.3269</b>	<b>0.2678</b>	<b>0.3270</b>	<b>0.3389</b>
1996	<b>0.1181</b>	<b>0.1890</b>	<b>0.1100</b>	<b>0.1132</b>	<b>0.1330</b>
1997	0.0315	<b>0.0581</b>	<b>0.0767</b>	<b>0.1299</b>	<b>0.0711</b>
1998	<b>0.2109</b>	<b>0.2984</b>	<b>0.4579</b>	<b>0.1065</b>	<b>0.1255</b>
1999	0.0357	<b>0.0758</b>	0.0393	0.0139	0.0234
2000	0.0261	<b>0.0562</b>	0.0292	<b>0.0806</b>	<b>0.0574</b>
2001	<b>0.5265</b>	<b>0.7227</b>	<b>0.7568</b>	<b>0.4987</b>	<b>0.5482</b>
2002	<b>0.1024</b>	<b>0.2074</b>	<b>0.0654</b>	<b>0.0964</b>	<b>0.1637</b>
2003	0.0232	<b>0.0833</b>	0.0386	<b>0.1303</b>	<b>0.1662</b>
2004	0.0041	0.0288	0.0040	0.0152	0.0243
2005	0.0251	<b>0.0891</b>	<b>0.0808</b>	0.0152	0.0185
2006	0.0080	0.0335	0.0282	0.0435	0.0186
2007	0.0255	<b>0.0707</b>	0.0238	0.0226	0.0206
2008	<b>0.1769</b>	<b>0.3234</b>	<b>0.2621</b>	<b>0.1357</b>	<b>0.1565</b>
2009	<b>0.4185</b>	<b>0.4381</b>	<b>0.4477</b>	<b>0.7069</b>	<b>0.6309</b>
2010	<b>0.0603</b>	<b>0.1401</b>	<b>0.0827</b>	<b>0.1110</b>	<b>0.1204</b>
2011	0.0455	<b>0.1098</b>	<b>0.0601</b>	0.0123	0.0123
2012	0.0211	<b>0.0577</b>	0.0127	0.0060	0.0096
2013	0.0106	<b>0.0358</b>	0.0019	0.0004	0.0019

Note: Abbreviations refer to AD: Anderson-Darling test (Anderson and Darling, 1952), CVM: Cramér-Von Mises test (Cramér, 1928; Mises, 1928), KS: Kolmogorov-Smirnov test (Massey, 1951), KUI: Kuiper test (Kuiper, 1960), and WU<sub>2</sub>: Watson U<sup>2</sup> test (Watson, 1961). P-values greater than 5 percent are shown in boldface. Entries equal to 0.0000 imply p-value  $< 5 \times 10^{-5}$ .

Table 19: Estimated parameters of the cross-sectional profit rate distribution.

In-sample period	$\hat{m}$	$\hat{\sigma}$
1980-1999	0.1024 (0.0007)	0.0564 (0.0006)
1981-2000	0.1010 (0.0006)	0.0556 (0.0006)
1982-2001	0.0992 (0.0007)	0.0552 (0.0006)
1983-2002	0.0976 (0.0007)	0.0548 (0.0005)
1984-2003	0.0956 (0.0007)	0.0544 (0.0005)
1985-2004	0.0935 (0.0007)	0.0539 (0.0006)
1986-2005	0.0926 (0.0006)	0.0537 (0.0005)
1987-2006	0.0920 (0.0006)	0.0533 (0.0005)
1988-2007	0.0914 (0.0006)	0.0534 (0.0006)
1989-2008	0.0908 (0.0006)	0.0539 (0.0006)
1990-2009	0.0895 (0.0006)	0.0543 (0.0006)
1991-2010	0.0887 (0.0007)	0.0545 (0.0006)

Note: Bootstrapped standard errors of the parameter estimates are shown in parentheses.

Inspired by this distributional regularity, Alfarano et al. (2012) construct the mean-reverting stochastic differential equation (SDE)

$$dX_t = -\frac{D}{2\sigma}\text{sign}(X_t - m)dt + \sqrt{D} dW_t \quad (49)$$

for the time evolution of the profit rate that has the Laplace in equation (48) as stationary density.<sup>49</sup> This diffusion process (henceforth AMIK process) is characterized by three parameters  $\theta = \{m, \sigma, D\}$ . It defines a diffusion on the real line around the measure of central tendency  $m \in \mathbb{R}$ , with dispersion  $\sigma \in \mathbb{R}^+$ . The parameter  $D \in \mathbb{R}^+$  is a diffusion coefficient and specifies the speed of mean-reversion, while  $dW_t$  represent Wiener increments. The process in equation (49) consists of two components. The diffusion function in the second term incorporates idiosyncratic shocks to profitability which are assumed to occur randomly with mean zero and variance  $D$ , while the drift function in the first term reflects the systematic tendency for competition to equalize profit rates across firms. Notice that the current realization of the profit rate determines merely the sign of the drift, not its strength, which in absolute terms is equal to  $D/(2\sigma)$ . As we shall see later, this constitutes a key difference to the competing Ornstein-Uhlenbeck process, for which the strength of the drift is proportional to the deviation from the average profit rate.

To estimate the diffusion process, I proceed as follows. Since equation (49) defines a Markov process with continuous trajectories, its transition probabilities obey the Fokker-Planck equation (see, e.g., Gardiner, 2009; Risken, 1996)

$$\frac{\partial p(x, t)}{\partial t} = -\frac{\partial}{\partial x}(A(x; \theta)p(x, t)) + \frac{1}{2} \frac{\partial^2}{\partial x^2}(B(x; \theta)p(x, t)), \quad (50)$$

<sup>49</sup> The distributional regularity in Figure 42 is obtained from *cross-sectional* data, while the dynamic law in equation (49) refers to the time-series. However, if (and only if) the system is in statistical equilibrium, cross-sectional and time-series properties will coincide so that the aggregate distributional regularity can be used to draw inferences on the time series of a single firm.



Table 20: Alternative specifications of the AMIK model.

Specification	Restricted parameter(s)	Free parameter(s)
M1		$m, \sigma, D$
M2	$m$	$\sigma, D$
M3	$\sigma$	$m, D$
M4	$m, \sigma$	$D$

where  $A(x; \theta)$  and  $B(x; \theta)$  are the drift and diffusion functions of the underlying process, and  $p(x, t) \equiv f(x, t | x_0, 0)$  denotes the conditional probability density for a transition from some initial state  $x_0$  at time 0 to state  $x$  at time  $t$ . The Fokker-Planck equation is a deterministic second-order partial differential equation that will be employed for ML estimation of the parameters and for obtaining the predictive density. Considering the initial condition of a unit probability mass in point  $x_0$ , the solution to the Fokker-Planck equation is

$$f(x, t | x_0, 0) = \frac{1}{\sqrt{2\pi Dt}} \cdot \exp \left\{ -\frac{(x - x_0)^2}{2Dt} - \frac{1}{2\sigma} (|x - m| - |x_0 - m|) - \frac{Dt}{8\sigma^2} \right\} + \frac{1}{2\sigma} \exp \left\{ -\frac{1}{\sigma} |x - m| \right\} \Phi \left( -\frac{|x - m| + |x_0 - m| - \frac{Dt}{2\sigma}}{\sqrt{Dt}} \right), \quad (51)$$

where  $\Phi(\cdot)$  is the cumulative distribution function of the standard normal distribution (Toda, 2012). Since there is no previous observation for  $x_0$ , I follow standard procedure in the field and evaluate this component using the unconditional probability density  $f_S(x)$  (see, e.g., Ghongadze and Lux, 2012; Lux, 2009). The log-likelihood of observations then amounts to

$$\log \mathcal{L}(\theta) = \log f_S(x_0; \theta) + \sum_{t=0}^{T-1} \log f(x_{t+1} | x_t; \theta). \quad (52)$$

It is maximized numerically to obtain the parameter estimates  $\hat{\theta} = \{\hat{m}, \hat{\sigma}, \hat{D}\}$ .

I distinguish four alternative model specifications in the estimation of the diffusion process in equation (49) which are summarized in Table 20. AMIK M1 is the most flexible model. It estimates all three parameters  $\{m, \sigma, D\}$  individually for each firm. Specification AMIK M2 has only two degrees of freedom since it assumes that all firms share a common location parameter, while  $\sigma$  and  $D$  remain firm-specific parameters. This assumption is motivated by the pre-analytical vision of a statistical equilibrium in firm profit rates. It implies that profit rates of all surviving firms in the sample regress toward the same value, which is the phenomenological  $\hat{m}$  from the cross-sectional profit rate distribution. In a similar vein, AMIK M3 estimates  $m$  and  $D$  from the time-series, while the dispersion parameter  $\sigma$  is estimated from the cross-sectional distribution, implying that the volatility of each time-series corresponds to the phenomenological  $\hat{\sigma}$  from the cross-sectional profit rate distribution. Finally, in

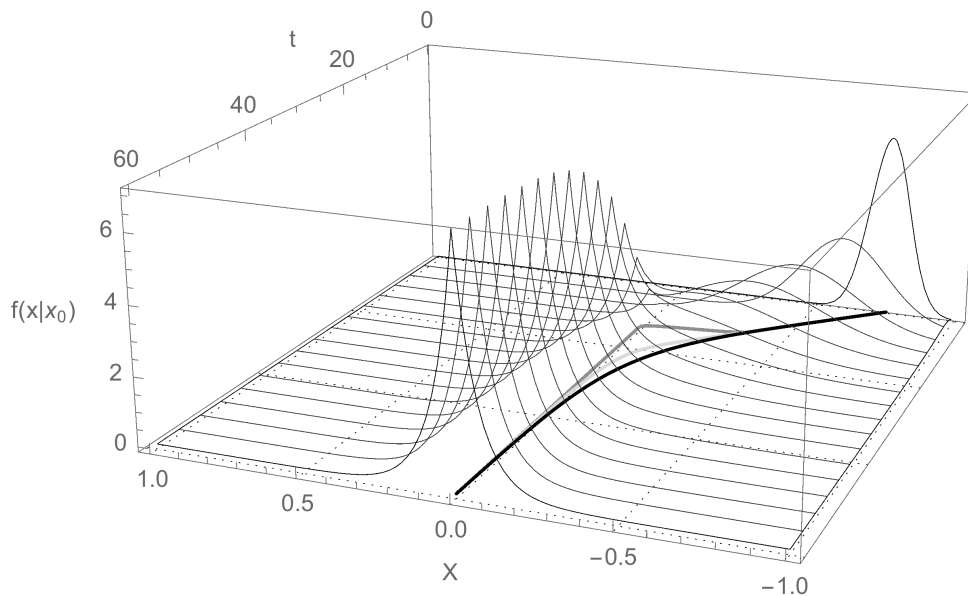


Figure 43: Evolution of the transient density of the AMIK diffusion conditional on the starting value  $x_0 = -0.8$ . The diffusion coefficient equals  $D = 0.005$ , and the standard deviation is  $\sqrt{2}\sigma = 0.1$ . As time increases the mode shifts toward  $m = 0$ . The three bold lines represent the mean (black), median (gray), and mode (light gray) of the transient density.

the AMIK M4 specification, I impose the restriction that both the location and scale parameter are identical to all firms and equal to their phenomenological values  $\hat{m}$  and  $\hat{\sigma}$ , respectively, implying that the profit rates of all firms revert to the same average and share a common volatility, as predicted by the statistical equilibrium methodology.

In addition to parameter estimation, the solution to the Fokker-Planck equation will be employed for obtaining the predictive density. As illustrated in Figure 43, we observe an asymmetric conditional probability density in the transient regime of the process, while it converges to the symmetric Laplace with unconditional mean  $m$  and long-term volatility  $\sigma$  for  $t \rightarrow \infty$ . Since forecast horizons up to 3 years are typically not long enough to reach this stationary distribution, one could consider the mean, mode, and median of the transient density as alternative predictors. Here, I will focus on the mean prediction because it yields the best forecasting performance. The expected value is determined by solving the integral

$$E[X] = \int_{-\infty}^{\infty} xf(x, t|x_0, 0) dx \quad (53)$$

by means of numerical integration.

### 4.3.2 Ornstein-Uhlenbeck process

A prominent alternative mean-reverting diffusion process is the model proposed by Uhlenbeck and Ornstein (1930):<sup>50</sup>

$$dX_t = \frac{D}{2\sigma^2}(m - X_t)dt + \sqrt{D}dW_t. \quad (54)$$

Unlike the AMIK process, the Ornstein-Uhlenbeck (OU) process has a drift term which depends linearly on the deviation from the unconditional mean  $m$ .<sup>51</sup> Put differently, the greater the distance between the actual realization and the long-term average, the stronger is the drift which pulls the profit rate to its unconditional mean. Thus, extreme profit rate realizations shall occur less frequently than in the AMIK model, leading to a more platykurtic distribution than the Laplace.<sup>52</sup>

Also for the OU process the corresponding Fokker-Planck equation can be solved analytically. Given the SDE in equation (54), the solution to equation (50) is

$$f(x, t|x_0, 0) = \frac{1}{\sqrt{2\pi\sigma^2(1 - \exp(-\frac{Dt}{\sigma^2}))}} \exp \left\{ -\frac{1}{2\sigma^2} \left( \frac{((x - m) - (x_0 - m)\exp(-\frac{Dt}{2\sigma^2}))^2}{1 - \exp(-\frac{Dt}{\sigma^2})} \right) \right\}, \quad (55)$$

which is Gaussian for all  $t$  with time-dependent first and second moment (see, e.g., Gardiner, 2009, p. 128). Figure 44 plots the transient density as a function of time. For  $t \rightarrow \infty$ , the conditional PDF converges to a stationary Normal distribution

$$f_S(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x - m)^2}{2\sigma^2} \right\} \quad (56)$$

with unconditional mean  $m$  and variance  $\sigma^2$ .

The OU process is estimated via ML using equations (52), (55) and (56). The forecast is the conditional mean of the transient density, given by

$$E[X] = x_0 \exp \left( -\frac{Dt}{2\sigma^2} \right) + m \left( 1 - \exp \left( -\frac{Dt}{2\sigma^2} \right) \right). \quad (57)$$

<sup>50</sup> Notice that the generic dispersion parameter  $\sigma$  refers to different volatility measures in equation (49) and equation (54). In the AMIK model,  $\sigma$  refers to the mean absolute deviation while it denotes the Gaussian standard deviation in the OU process. Yet one can compute one measure from the other since a Laplace-distributed random variable  $X \sim L(m, \sigma)$  has standard deviation  $\sqrt{2}\sigma$ .

<sup>51</sup> Another diffusion process that exhibits mean-reverting behavior is the model by Cox et al. (1985) which is frequently used for the modeling of interest rates. In this model the conditional volatility of random innovations in the diffusion function prevents negative realizations of the process. Since a zero lower bound is absent for the profit rate, I do not consider this model here.

<sup>52</sup> Since the OU process is the continuous time analog of a stationary AR(1) process, it should be similar to the model used in Fairfield et al. (2009). However, their model includes the predicted value of sales growth as an additional variable.

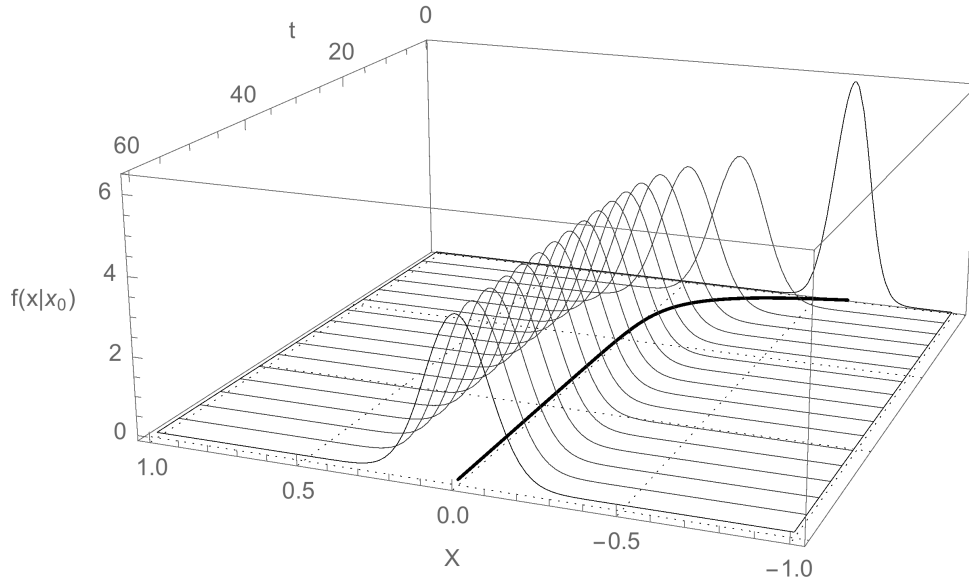


Figure 44: Evolution of the transient density of the OU process conditional on the starting value  $x_0 = -0.8$ . The diffusion coefficient equals  $D = 0.005$  and the standard deviation is  $\sigma = 0.1$ . As time increases the mode shifts towards  $m = 0$ . The bold black line represents the mean prediction.

#### 4.3.3 *AR(I)MA-type models*

As additional candidates, I consider time-series models from the mixed autoregressive and moving average varieties. Despite their simplicity, these models may capture some of the aforementioned data characteristics for sensible choices of the parameters, in particular mean-reversion and positive autocorrelations. Here, I will consider  $ARMA(p, q)$  as well as more general  $ARIMA(p, d, q)$  models.

The  $ARMA(p, q)$  model reads

$$\left(1 - \sum_{i=1}^p \lambda_i B^i\right) X_t = c + \left(1 + \sum_{i=1}^q \psi_i B^i\right) \varepsilon_t, \quad (58)$$

where  $B$  represents the backshift operator,  $\Lambda(B) = 1 - \lambda_1 B - \lambda_2 B^2 - \dots - \lambda_p B^p$  and  $\Psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots + \psi_q B^q$  are the autoregressive and moving average polynomials, respectively, and  $\varepsilon_t$  denotes a white noise series with  $E[\varepsilon_t] = 0$ ,  $E[\varepsilon_t^2] = \sigma^2$  and  $E[\varepsilon_t \varepsilon_\tau] = 0$  for  $t \neq \tau$ .  $c$  represents a constant term that is included to capture a possibly non-zero mean of the process.

I estimate the  $p + q + 2$  model parameters in  $\Xi = (c, \lambda_1, \lambda_2, \dots, \lambda_p, \psi_1, \psi_2, \dots, \psi_q, \sigma^2)'$  for  $1 \leq p, q \leq 5$  via ML assuming that  $\varepsilon_t$  are i.i.d. Gaussian random variables (see, e.g., Hamilton, 1994, ch. 5 for details). From the set of estimated models within this range I

choose the specification that minimizes the Schwarz (1978) information criterion (SIC). As it turns out, the parsimonious, yet effective, ARMA(1,1) specification

$$X_t = c + \lambda X_{t-1} + \epsilon_t + \psi \epsilon_{t-1} \quad (59)$$

comes out as the most favored model across firms, and is therefore chosen for forecasting.

To predict with ARMA, I follow the standard procedure for obtaining linear forecasts, as outlined, e.g., in Diebold (2006), ch. 8. First, the process is updated in the time dimension to period  $t + h$ . Then, I project on the information set available in  $t$ , the period when the forecast is made, which contains current and past realizations of the response variable and innovations. To this end, all future values of the response variable and innovations are replaced with their forecasts. The latter are zero by definition, while the former are obtained via the recursive method (also known as chain rule of forecasting). Thus, the  $h$ -step ahead forecast for an ARMA(1,1) process conditional on the information set available in  $t$  is

$$\hat{x}_{t+h} = \begin{cases} \hat{c} + \lambda x_t + \psi \epsilon_t & \text{if } h = 1, \\ \hat{c} + \lambda \hat{x}_{t+h-1} & \text{if } h > 1. \end{cases} \quad (60)$$

In addition to ARMA( $p, q$ ) processes, I have also estimated mixed autoregressive integrated moving average models

$$\left(1 - \sum_{i=1}^p \lambda_i B^i\right) (1 - B)^d X_t = \left(1 + \sum_{i=1}^q \psi_i B^i\right) \epsilon_t \quad (61)$$

to account for a potential non-stationarity of the data. For  $1 \leq p, q \leq 5$  and  $d \geq 1$ , ARIMA(1,1,1) compares most favorably based on SIC. Thus, I fit an ARMA(1,1) model to the first difference of the original series, which is then used for forecasting.<sup>53</sup> Predictions for the original data can be obtained by successively adding the forecast of the first difference (i.e. the change in that variable) to the last observation of the original data,  $x$ , or its forecast,  $\hat{x}$ , respectively.

Finally, I complete the set of competitors by considering the random walk (RW) without drift

$$X_t = X_{t-1} + \epsilon_t, \quad (62)$$

which is a special case of ARMA(1,0) with unit root. Obviously, in case of the RW no parameters need to be estimated and the best prediction is the naïve forecast

$$\hat{x}_{t+h} = x_t \quad \forall h > 0. \quad (63)$$

---

<sup>53</sup> Here I drop the constant because it determines the drift (not the mean) in an ARIMA model.

Table 21: Summary of forecasting performances: number of winnings in the test for SPA.

h	MSE					MAE				
	Benchmark					Benchmark				
	AMIK	OU	ARMA	ARIMA	RW	AMIK	OU	ARMA	ARIMA	RW
	M <sub>1</sub>					M <sub>1</sub>				
1	404	419	278	190	<b>459</b>	414	422	276	181	<b>461</b>
2	363	368	297	163	<b>386</b>	365	365	292	166	<b>400</b>
3	341	336	290	128	<b>348</b>	355	332	296	129	<b>357</b>
	AMIK	OU	ARMA	ARIMA	RW	AMIK	OU	ARMA	ARIMA	RW
	M <sub>2</sub>					M <sub>2</sub>				
1	<b>443</b>	416	280	188	428	<b>435</b>	413	276	175	428
2	<b>392</b>	378	297	162	339	<b>404</b>	379	293	161	363
3	<b>370</b>	342	283	115	284	<b>382</b>	344	304	115	309
	AMIK	OU	ARMA	ARIMA	RW	AMIK	OU	ARMA	ARIMA	RW
	M <sub>3</sub>					M <sub>3</sub>				
1	<b>451</b>	418	279	188	434	<b>447</b>	415	281	170	432
2	<b>407</b>	378	305	160	368	<b>403</b>	382	295	154	378
3	<b>384</b>	355	298	125	335	<b>388</b>	346	309	128	345
	AMIK	OU	ARMA	ARIMA	RW	AMIK	OU	ARMA	ARIMA	RW
	M <sub>4</sub>					M <sub>4</sub>				
1	<b>456</b>	413	284	190	419	<b>435</b>	416	270	176	434
2	<b>405</b>	369	279	158	328	<b>407</b>	378	290	160	351
3	<b>380</b>	342	287	119	281	<b>386</b>	344	304	120	301

Note: The null hypothesis is that the benchmark model is not outperformed by other models in terms of MSE and MAE, respectively. Entries refer to the number of firms for which the null hypothesis cannot be rejected on the 10 percent level. For each loss function the highest number of winnings is shown in boldface.

#### 4.4 RESULTS

Tables 22-28 in Appendix D.2 show summary statistics for the estimated parameters of AMIK M<sub>1</sub>-M<sub>4</sub>, OU, ARMA, and ARIMA models. For ARMA it turns out that the estimated autoregressive parameters are smaller than unity in absolute value in 478-498 out of 516 cases (depending on the period under consideration), indicating stationarity of the vast majority of fitted models. Based on the estimation results, I construct out-of-sample forecasts and compute squared and absolute forecast errors for comparison of forecast accuracy. Table 21 presents the SPA test results for different benchmark models and forecast horizons up to three years ahead. From these I infer the following regularities:

1. Perhaps somewhat surprisingly, RW is not outperformed by the unrestricted AMIK M<sub>1</sub> process and its mean-reverting competitors in most cases. This holds particularly for shorter forecast horizons,  $h \leq 2$ , for which the difference between RW and the other models is most pronounced. For  $h = 3$ , RW and the second

best AMIK M<sub>1</sub> achieve nearly identical results, implying that the random walk's initial advantage is averaged out over longer time horizons. Nevertheless, the respectable performance of RW suggests that the profit rate exhibits considerable random fluctuations and that models predicting mean-reversion to firm-specific profitability levels are not superior to the naïve forecast.

2. AMIK M<sub>1</sub> and OU yield quite similar performances. For one year ahead forecasts, the latter achieves the highest number of winnings. Yet the difference to AMIK is small and probably not due to a systematic advantage. For longer forecast horizons, however, differences between the two models become more pronounced and turn in favor of AMIK, in particular for the MAE criterion.
3. Turning to the Box-Jenkins models, ARMA is clearly dominated by AMIK, OU, and RW for both the MSE and MAE criterion. Moreover, there is no indication that differencing improves the forecasting capacity since ARIMA performs consistently worse than ARMA. Although the information criterion clearly favors  $d = 1$  in ARIMA( $p, d, q$ ), I have also experimented with higher order differencing, yet ARIMA remains inferior to the other models also for these specifications. Overall, the bad performance of ARIMA models is in line with the results reported in Section 4.2 which already suggest that non-stationarity of the data is (if at all) a minor issue, although the analysis builds on raw data.
4. Considering the full set of AMIK specifications M<sub>1</sub> to M<sub>4</sub>, I find support for the statistical equilibrium methodology since the relative forecasting performance of AMIK improves if the phenomenological values of  $m$  and  $\sigma$  replace idiosyncratic location and dispersion parameters. In case of the MSE (MAE) criterion, AMIK M<sub>4</sub> yields 37 (1), 36 (29), and 38 (42) more winnings than the second best model for one, two, and three years ahead forecasts, respectively, while the difference between AMIK M<sub>4</sub> and the naïve RW forecast amounts to 37 (1), 77 (56), and 99 (85). This implies that RW can only be outperformed if it is assumed that the system is in statistical equilibrium.
5. Comparing the models with two degrees of freedom, AMIK M<sub>2</sub> and M<sub>3</sub>, it appears that fixing the dispersion parameter  $\sigma$  leads to more forecasting gains than fixing the location parameter  $m$ . I infer that this is due to the short testing sample since previous investigations indicate that the median half life of abnormal profitability is about 9 years (Mundt et al., 2016). Thus, most firms should need more than 3 years to converge to the system-wide average profit rate.

Two insights from additional experiments may be worth reporting as well:

6. Since the transient density of the AMIK diffusion becomes asymmetric for some  $t$ , I have also experimented with alternatives to the expected value. However, it turns out that both the median and the mode of the conditional probability density function are clearly dominated by the expected value, i.e. their SPA test results are inferior to the mean prediction (results not shown). The failure of the

mode should arise from the abrupt jump of the transient density's maximum that can be observed in Figure 43.

7. I also computed MSE and MAE loss functions by averaging forecasting errors across firms to obtain an overall assessment of each model's relative forecasting performance. An advantage of this approach is that the number of observations in the test for SPA rises from 12 (number of time shifts of the rolling window) to 516 (number of firms in the sample).<sup>54</sup> While the test indicates that ARMA and ARIMA models are inferior to their competitors, the p-values of the remaining models are typically so high that there are no statistically significant differences between them at the usual confidence levels.<sup>55</sup> This confirms the results reported in Table 21 which already suggest that the forecasting performances of AMIK, OU, and RW are relatively similar and that significant differences between the models occur only in approximately 40-100 out of 516 cases.

Finally, since the notion of ergodicity does not relate to any specific type of distribution, say, Laplace or Gaussian, I also considered to what extent the performance of the other (stationary) models might improve if firm-level parameter estimates are replaced with ensemble averages. For OU, I follow the procedure outlined in Section 4.3.1. The only difference from AMIK is that the cross-sectional parameters are now estimated from the PDF in equation (56) because the stationary distribution of OU is Gaussian. Adding restrictions to the ARMA model is a little more cumbersome. Since the mean  $m$  of a stationary ARMA(1,1) process is given by

$$m = \frac{c}{1 - \lambda}, \quad (64)$$

we can express the constant in terms of  $c = m(1 - \lambda)$  and substitute the right hand side of this expression for  $c$  in equation (59). This yields

$$X_t - m = \lambda(X_{t-1} - m) + \epsilon_t + \psi\epsilon_{t-1}. \quad (65)$$

Thus, we obtain a restricted model by fitting an ARMA process to the centered time series. The latter is obtained by subtracting the cross-sectional sample mean from the time series. I also tried to fix the variance, but estimation of the restricted model yields many cases (firms) for which  $\lambda$  is close to  $-\psi$ , indicating an overparametrization of the model that should be avoided (see Hamilton, 1994, p. 60).<sup>56</sup> I take this to imply that it is not readily feasible to fix the variance of this model, and hence do not report results for ARMA M<sub>3</sub> and ARMA M<sub>4</sub>. Table 29 in Appendix D.3 reports the corresponding SPA test results for the remaining (restricted) models. I find that AMIK M<sub>2</sub> and M<sub>3</sub> are still superior to the alternatives for forecast horizons  $h \geq 2$  years, while superiority of AMIK M<sub>4</sub> against OU M<sub>4</sub> prevails only for the longest forecast horizon  $h = 3$ .

<sup>54</sup> Here, I employed a fixed scheme instead of a rolling window scheme.

<sup>55</sup> This material is available upon request.

<sup>56</sup> In this case, a simple white noise process  $X_t = \epsilon_t$  would model the data equally well because  $(1 - \lambda B) \approx (1 + \psi B)$  in  $(1 - \lambda B)X_t = (1 + \psi B)\epsilon_t$ .



Overall, these results strongly suggest that parameter estimation from cross-sectional data improves the forecasting performance of all models, and that (under identical conditions) the correctly specified AMIK process outperforms the misspecified OU model only for longer time horizons.<sup>57</sup>

#### 4.5 CONCLUDING REMARKS

In this paper, I employ the statistical equilibrium model by Alfarano and Milaković (2008) and Alfarano et al. (2012) to forecast profitability of long-lived corporations. The model is consistent with the leptokurtic profit rate density that is well approximated by a symmetric Laplace distribution and the autocorrelation structure of the data. Moreover, in line with previous theoretical (e.g. Ohlson and Juettner-Nauroth, 2005) and empirical studies (e.g. Fairfield et al., 2009), it accounts for the idea that profitability reverts to an economy-wide average because the notion of statistical equilibrium implies that profit rates are drawn from the same distribution. While pure autoregressive or mixed ARMA-type models can at best account for persistence in profitability and mean-reversion to economy-wide profitability levels, they are not consistent with the Laplacian nature of profit rates and are, therefore, counterfactual. Thus, to the best of my knowledge, the new approach is the only model that simultaneously accounts for all these characteristics of the data.

A general problem of forecasting firm profitability is the relatively small amount of accounting data that is available for model estimation and forecast evaluation, which could impair the detection of significant differences between the models. Although the present data set is certainly not ideal from a statistical viewpoint, the results are encouraging in that they indicate weak superiority of the statistical equilibrium model vis-à-vis its competitors, particularly for longer forecast horizons. One of the main insights is that the assumption of ergodicity is absolutely essential to realize forecasting gains relative to the random walk, which turns out to be a tough competitor for standard time series and more structural diffusion models, as it has been reported by Meese and Rogoff (1983a) and Meese and Rogoff (1983b) in the context of forecasting exchange rates.

While one could have hoped to obtain more definite results, it is not even clear whether the AMIK diffusion would show better forecasting performance if it was the “true” data generating process, mainly because of random innovations in the diffusion function. These innovations capture idiosyncratic effects on profitability arising from complex interactions among firms, yet the stochasticity of these shocks prevents accurate prediction of the profit rate. In a sense, the respectable performance of the random walk testifies to these fluctuations that prevail even in (statistical) equilibrium.<sup>58</sup> Forecasting becomes even more cumbersome because the diffusion coefficient

---

<sup>57</sup> OU is misspecified in the sense that its stationary distribution is Gaussian, whereas the empirical density of the data is approximately Laplace.

<sup>58</sup> Notice that statistical equilibrium refers to a stationary *distribution* of states within the system, not to an equilibrium state of every single individual. In fact, the position of any two individuals in that equilibrium distribution is perfectly interchangeable.

$D$  appears in both the drift and the diffusion function of the AMIK model. On the one hand, a larger  $D$  would imply faster mean-reversion which should improve the performance of AMIK relative to its competitors as differences in autocorrelations become more pronounced. However, a larger diffusion coefficient also implies a larger variance of idiosyncratic shocks to profitability. On the other hand, if  $D$  is small, a low noise level comes along with slow mean-reversion, so that the alternative (and probably misspecified) models might represent a reasonable approximation to the AMIK diffusion in the short run.<sup>59</sup>

Several of the arguments raised before already suggest that there are limitations to predicting individual behavior when the environment is characterized by intense interactions among individuals. As I have repeatedly emphasized, the statistical equilibrium model explicitly accounts for this observation by stating that profit rates are governed by both the systematic drift towards the system-wide average *and* random fluctuations. In such an environment, forecasting performance may not be the most telling criterion for goodness of fit because individual behaviors are affected by so many things that they are, essentially, extremely hard to predict. Overall, my finding that, under identical conditions, the misspecified OU process and the correctly specified AMIK process yield very similar performances indicates that forecasting accuracy is an extremely fragile criterion. What should receive more attention are robust aggregate properties that arise from these interactions among firms, i.e. the cross-sectional distribution of profit rates, which constitute the backbone of this statistical equilibrium approach.

---

<sup>59</sup> Ghongadze and Lux (2012) observe a similar problem for a stochastic model of economic sentiment indicators.

## APPENDIX

## D.1 DESCRIPTIVE STATISTICS

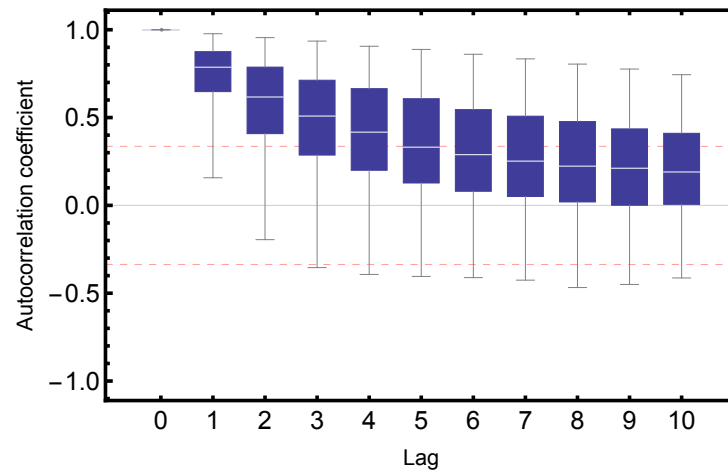


Figure 45: Box-Whisker chart for the estimated autocorrelation functions of profit rates. The boxes include the 25, 50, and 75 percent quantiles. 95 percent asymptotic confidence intervals are shown as red dashed lines.

## D.2 ESTIMATION RESULTS

Table 22: Summary statistics for estimated parameters of AMIK M<sub>1</sub>.

In-sample period	Min.	25%	Median	75%	Max.
$\hat{m}$					
1980-1999	-0.1630	0.0714	0.0987	0.1417	0.4994
1981-2000	-0.1630	0.0701	0.0978	0.1392	0.3254
1982-2001	-0.1413	0.0680	0.0966	0.1330	0.5698
1983-2002	-0.1413	0.0668	0.0943	0.1304	0.4478
1984-2003	-0.1413	0.0660	0.0932	0.1280	0.4429
1985-2004	-0.1413	0.0648	0.0906	0.1290	0.4429
1986-2005	-0.1413	0.0668	0.0903	0.1246	0.2682
1987-2006	-0.1630	0.0669	0.0902	0.1246	0.3795
1988-2007	-0.1655	0.0673	0.0900	0.1220	0.4346
1989-2008	-0.2043	0.0658	0.0906	0.1220	0.3795
1990-2009	-0.2043	0.0594	0.0857	0.1173	0.3795
1991-2010	-0.1655	0.0603	0.0864	0.1165	0.4145
$\hat{\sigma}$					
1980-1999	0.0044	0.0202	0.0312	0.0471	0.2356
1981-2000	0.0039	0.0198	0.0316	0.0467	0.2170
1982-2001	0.0049	0.0197	0.0326	0.0468	0.1905
1983-2002	0.0039	0.0203	0.0319	0.0458	0.1885
1984-2003	0.0039	0.0202	0.0311	0.0458	0.1845
1985-2004	0.0040	0.0196	0.0311	0.0450	0.1936
1986-2005	0.0036	0.0198	0.0309	0.0450	0.1581
1987-2006	0.0032	0.0197	0.0303	0.0450	0.1812
1988-2007	0.0024	0.0193	0.0305	0.0455	0.2059
1989-2008	0.0035	0.0200	0.0304	0.0477	0.2213
1990-2009	0.0025	0.0196	0.0309	0.0478	0.2220
1991-2010	0.0024	0.0188	0.0307	0.0481	0.2552
$\hat{D}$					
1980-1999	0.0000	0.0005	0.0013	0.0038	4.3628
1981-2000	0.0000	0.0005	0.0014	0.0039	1.0983
1982-2001	0.0000	0.0005	0.0014	0.0036	1.0826
1983-2002	0.0000	0.0005	0.0014	0.0033	2.0623
1984-2003	0.0000	0.0005	0.0012	0.0030	1.5877
1985-2004	0.0000	0.0005	0.0011	0.0028	7.7582
1986-2005	0.0000	0.0004	0.0011	0.0029	3.7175
1987-2006	0.0000	0.0004	0.0010	0.0026	11.5684
1988-2007	0.0000	0.0004	0.0010	0.0024	3.7328
1989-2008	0.0000	0.0004	0.0010	0.0027	7.2643
1990-2009	0.0000	0.0004	0.0011	0.0028	12.8265
1991-2010	0.0000	0.0004	0.0011	0.0028	4.6466

Note: Entries equal to 0.0000 imply that the corresponding values are smaller than  $5 \times 10^{-5}$ .

Table 23: Summary statistics for estimated parameters of AMIK M2.

In-sample period	Min.	25%	Median	75%	Max.
$\hat{\sigma}$					
1980-1999	0.0087	0.0408	0.0610	0.0916	0.5061
1981-2000	0.0089	0.0402	0.0611	0.0897	0.4425
1982-2001	0.0078	0.0403	0.0617	0.0905	0.5569
1983-2002	0.0074	0.0402	0.0599	0.0913	0.4954
1984-2003	0.0078	0.0389	0.0578	0.0910	0.4894
1985-2004	0.0087	0.0368	0.0558	0.0883	0.6204
1986-2005	0.0097	0.0360	0.0567	0.0849	0.4524
1987-2006	0.0095	0.0354	0.0555	0.0851	0.4983
1988-2007	0.0098	0.0346	0.0557	0.0867	0.6071
1989-2008	0.0097	0.0357	0.0574	0.0894	0.6285
1990-2009	0.0105	0.0343	0.0588	0.0844	0.6126
1991-2010	0.0103	0.0338	0.0565	0.0828	0.6013
$\hat{D}$					
1980-1999	0.0000	0.0004	0.0011	0.0029	9.5482
1981-2000	0.0000	0.0004	0.0011	0.0030	24.7156
1982-2001	0.0000	0.0004	0.0011	0.0029	12.5131
1983-2002	0.0000	0.0004	0.0010	0.0026	63.9495
1984-2003	0.0000	0.0004	0.0010	0.0025	31.2953
1985-2004	0.0000	0.0004	0.0009	0.0025	12.8177
1986-2005	0.0000	0.0004	0.0009	0.0024	28.0575
1987-2006	0.0000	0.0004	0.0009	0.0023	17.1738
1988-2007	0.0000	0.0004	0.0009	0.0022	24.4018
1989-2008	0.0000	0.0003	0.0009	0.0022	30.4482
1990-2009	0.0000	0.0004	0.0009	0.0025	65.9330
1991-2010	0.0000	0.0004	0.0009	0.0024	10.2520

*Note:* The location parameter  $m$  is fixed at its phenomenological value. Entries equal to 0.0000 imply that the corresponding values are smaller than  $5 \times 10^{-5}$ .

Table 24: Summary statistics for estimated parameters of AMIK  $M_3$ .

In-sample period	Min.	25%	Median	75%	Max.
$\hat{m}$					
1980-1999	-0.1630	0.0715	0.0992	0.1418	0.2979
1981-2000	-0.1630	0.0696	0.0967	0.1366	0.3254
1982-2001	-0.1413	0.0684	0.0968	0.1331	0.5698
1983-2002	-0.1413	0.0668	0.0941	0.1304	0.4478
1984-2003	-0.1630	0.0659	0.0929	0.1280	0.4429
1985-2004	-0.1413	0.0655	0.0909	0.1280	0.4429
1986-2005	-0.1413	0.0667	0.0904	0.1246	0.2682
1987-2006	-0.1630	0.0671	0.0904	0.1265	0.4145
1988-2007	-0.1655	0.0667	0.0895	0.1216	0.4346
1989-2008	-0.2043	0.0658	0.0902	0.1208	0.3795
1990-2009	-0.2043	0.0604	0.0858	0.1179	0.3790
1991-2010	-0.1661	0.0598	0.0864	0.1173	0.4145
$\hat{D}$					
1980-1999	0.0000	0.0004	0.0011	0.0032	4.5497
1981-2000	0.0000	0.0005	0.0011	0.0032	5.1624
1982-2001	0.0000	0.0005	0.0012	0.0029	13.6392
1983-2002	0.0000	0.0005	0.0011	0.0026	3.7498
1984-2003	0.0000	0.0005	0.0011	0.0025	2.1020
1985-2004	0.0000	0.0004	0.0010	0.0025	3.6174
1986-2005	0.0000	0.0004	0.0010	0.0025	4.0469
1987-2006	0.0000	0.0004	0.0010	0.0023	1.2582
1988-2007	0.0000	0.0004	0.0009	0.0023	13.6962
1989-2008	0.0000	0.0004	0.0009	0.0024	3.7846
1990-2009	0.0000	0.0004	0.0010	0.0026	28.3821
1991-2010	0.0000	0.0004	0.0010	0.0025	3.0516

Note: The dispersion parameter  $\sigma$  is fixed at its phenomenological value. Entries equal to 0.0000 imply that the corresponding values are smaller than  $5 \times 10^{-5}$ .

Table 25: Summary statistics for estimated parameters of AMIK M4.

In-sample period	Min.	25%	Median	75%	Max.
	$\hat{D}$				
1980-1999	0.0000	0.0004	0.0010	0.0027	9.6918
1981-2000	0.0000	0.0004	0.0010	0.0026	6.9888
1982-2001	0.0000	0.0004	0.0011	0.0026	0.3425
1983-2002	0.0000	0.0004	0.0010	0.0025	0.1601
1984-2003	0.0000	0.0004	0.0009	0.0023	2.2816
1985-2004	0.0000	0.0004	0.0009	0.0023	1.5836
1986-2005	0.0000	0.0004	0.0009	0.0023	1.5836
1987-2006	0.0000	0.0004	0.0008	0.0021	1.8736
1988-2007	0.0000	0.0004	0.0008	0.0021	12.5901
1989-2008	0.0000	0.0003	0.0008	0.0022	9.6918
1990-2009	0.0000	0.0003	0.0009	0.0025	1.3057
1991-2010	0.0000	0.0003	0.0009	0.0023	9.6918

*Note:* The location parameter  $m$  and the dispersion parameter  $\sigma$  are fixed at their phenomenological values. Entries equal to 0.0000 imply that the corresponding values are smaller than  $5 \times 10^{-5}$ .

Table 26: Summary statistics for estimated parameters of the OU process.

In-sample period	Min.	25%	Median	75%	Max.
$\hat{\mu}$					
1980-1999	-0.0514	0.0736	0.1037	0.1400	0.4210
1981-2000	-0.0649	0.0717	0.1010	0.1382	0.3799
1982-2001	-0.0345	0.0689	0.0974	0.1317	0.4038
1983-2002	-0.1391	0.0693	0.0968	0.1315	0.4265
1984-2003	-0.1586	0.0673	0.0957	0.1295	0.4017
1985-2004	-0.1475	0.0674	0.0940	0.1263	0.3954
1986-2005	-0.1935	0.0664	0.0930	0.1270	0.3727
1987-2006	-0.2091	0.0685	0.0932	0.1235	0.4156
1988-2007	-0.2710	0.0672	0.0917	0.1236	0.4088
1989-2008	-0.7252	0.0662	0.0906	0.1222	0.3997
1990-2009	-0.2813	0.0622	0.0873	0.1172	0.4076
1991-2010	-0.3333	0.0627	0.0865	0.1182	0.4295
$\hat{\sigma}$					
1980-1999	0.0071	0.0261	0.0377	0.0599	0.4101
1981-2000	0.0062	0.0248	0.0381	0.0578	0.4267
1982-2001	0.0052	0.0246	0.0395	0.0574	0.2580
1983-2002	0.0046	0.0251	0.0373	0.0558	0.2989
1984-2003	0.0060	0.0249	0.0381	0.0576	0.2587
1985-2004	0.0060	0.0243	0.0369	0.0584	0.2374
1986-2005	0.0056	0.0247	0.0371	0.0583	0.2520
1987-2006	0.0047	0.0242	0.0365	0.0577	0.2854
1988-2007	0.0056	0.0242	0.0367	0.0576	0.3695
1989-2008	0.0040	0.0243	0.0372	0.0580	0.9164
1990-2009	0.0048	0.0239	0.0370	0.0583	0.5244
1991-2010	0.0038	0.0233	0.0374	0.0566	0.5416
$\hat{D}$					
1980-1999	0.0000	0.0005	0.0014	0.0037	0.4382
1981-2000	0.0000	0.0005	0.0014	0.0038	6.9470
1982-2001	0.0000	0.0005	0.0013	0.0036	3.6951
1983-2002	0.0000	0.0005	0.0014	0.0031	0.4925
1984-2003	0.0000	0.0005	0.0012	0.0029	0.1651
1985-2004	0.0000	0.0005	0.0011	0.0027	3.0114
1986-2005	0.0000	0.0005	0.0011	0.0028	0.3227
1987-2006	0.0000	0.0004	0.0011	0.0026	1.6097
1988-2007	0.0000	0.0004	0.0010	0.0025	0.3960
1989-2008	0.0000	0.0004	0.0010	0.0027	0.9644
1990-2009	0.0000	0.0004	0.0011	0.0029	0.3035
1991-2010	0.0000	0.0004	0.0011	0.0028	4.1833

Note: Entries equal to 0.0000 imply that the corresponding values are smaller than  $5 \times 10^{-5}$ .



Table 27: Summary statistics for estimated parameters of the ARMA model.

In-sample period	Min.	25%	Median	75%	Max.
$\hat{c}$					
1980-1999	-0.9227	0.0272	0.0515	0.0845	0.9339
1981-2000	-104.7131	0.0265	0.0481	0.0827	4.6861
1982-2001	-10.4311	0.0275	0.0481	0.0821	1.0418
1983-2002	-122.2397	0.0267	0.0484	0.0834	0.5198
1984-2003	-1.2884	0.0259	0.0473	0.0802	0.9154
1985-2004	-2.9854	0.0261	0.0455	0.0754	7.9124
1986-2005	-2.6578	0.0256	0.0426	0.0731	2.5900
1987-2006	-0.5794	0.0237	0.0431	0.0697	0.8173
1988-2007	-1.1162	0.0232	0.0418	0.0676	9.0449
1989-2008	-7.0722	0.0224	0.0394	0.0672	1.8875
1990-2009	-1.9883	0.0220	0.0403	0.0664	11.1863
1991-2010	-2.3408	0.0212	0.0404	0.0671	9.9020
$\hat{\lambda}$					
1980-1999	-42.8141	0.1498	0.5025	0.6903	10.9541
1981-2000	-76.4585	0.1493	0.5061	0.7058	1182.9813
1982-2001	-7.5930	0.0999	0.4974	0.6975	156.6820
1983-2002	-425.9735	0.1319	0.4942	0.6721	2102.9842
1984-2003	-7.8685	0.1605	0.4893	0.7042	15.0287
1985-2004	-75.0235	0.1978	0.5124	0.6854	46.9542
1986-2005	-33.0459	0.2100	0.5219	0.6888	18.0523
1987-2006	-8.5463	0.2636	0.5350	0.7097	8.2472
1988-2007	-96.8263	0.2857	0.5296	0.7100	12.7770
1989-2008	-18.6911	0.2915	0.5438	0.7165	108.8221
1990-2009	-148.6684	0.2922	0.5510	0.7193	22.6550
1991-2010	-216.5744	0.2827	0.5481	0.7182	42.7274
$\hat{\psi}$					
1980-1999	-14.9776	-0.0538	0.2002	0.5437	42.8123
1981-2000	-1182.9816	-0.0676	0.1968	0.5515	76.4629
1982-2001	-156.6805	-0.0626	0.2140	0.5806	7.6365
1983-2002	-2102.9835	-0.0107	0.2584	0.5556	425.9736
1984-2003	-15.0498	-0.0419	0.1928	0.5137	7.9088
1985-2004	-46.9609	-0.0229	0.1945	0.4754	75.0284
1986-2005	-18.0662	-0.0404	0.2149	0.5069	33.0370
1987-2006	-8.2253	-0.0300	0.2135	0.5087	8.5292
1988-2007	-12.8117	-0.0258	0.2049	0.4988	96.8289
1989-2008	-108.8210	-0.0226	0.1703	0.4757	18.7106
1990-2009	-22.6645	-0.0697	0.1834	0.4915	148.6667
1991-2010	-42.7190	-0.0446	0.1903	0.4979	216.5742
$\hat{\sigma}^2$					
1980-1999	0.0000	0.0004	0.0008	0.0019	0.0843
1981-2000	0.0000	0.0004	0.0008	0.0020	0.0769
1982-2001	0.0000	0.0004	0.0009	0.0019	0.0426
1983-2002	0.0000	0.0004	0.0008	0.0017	0.0454
1984-2003	0.0000	0.0004	0.0008	0.0017	0.0353
1985-2004	0.0000	0.0003	0.0008	0.0017	0.0352
1986-2005	0.0000	0.0003	0.0008	0.0017	0.0303
1987-2006	0.0000	0.0003	0.0007	0.0016	0.0348
1988-2007	0.0000	0.0003	0.0007	0.0016	0.0297
1989-2008	0.0000	0.0003	0.0007	0.0016	0.2208
1990-2009	0.0000	0.0003	0.0008	0.0018	0.1902
1991-2010	0.0000	0.0003	0.0008	0.0016	0.1796

Note: Entries equal to 0.0000 imply that the corresponding values are smaller than  $5 \times 10^{-5}$ .

Table 28: Summary statistics for estimated parameters of the ARIMA model.

In-sample period	Min.	25%	Median	75%	Max.
$\hat{\lambda}$					
1980-1999	-211.0938	-0.5234	0.1014	0.7645	174.8884
1981-2000	-80.4232	-0.5979	0.1289	0.7736	22.4037
1982-2001	-160.4340	-0.5599	0.1647	0.7962	1089.6294
1983-2002	-353.2678	-0.4222	0.2198	0.8769	110.0544
1984-2003	-47154.0926	-0.5238	0.1389	0.7083	1054.2162
1985-2004	-108.2522	-0.5494	0.1340	0.7729	51.5886
1986-2005	-68.8545	-0.5947	0.0721	0.7039	46.1794
1987-2006	-71.2947	-0.5500	0.0640	0.6925	174.8354
1988-2007	-68.6584	-0.5735	0.0575	0.7256	149.5186
1989-2008	-2892.6316	-0.5915	0.0843	0.7117	612.8428
1990-2009	-99.5175	-0.5716	0.1753	0.7330	223.8485
1991-2010	-51.6857	-0.4959	0.1667	0.6779	419.4589
$\hat{\psi}$					
1980-1999	-174.8887	-1.0020	-0.4312	0.5400	211.0947
1981-2000	-22.3862	-1.0006	-0.4645	0.6494	80.4274
1982-2001	-1089.6297	-1.0011	-0.5632	0.4500	160.4363
1983-2002	-110.0584	-1.1239	-0.6171	0.2780	353.2683
1984-2003	-1054.2164	-1.0000	-0.4398	0.4481	47154.0926
1985-2004	-51.5841	-1.0120	-0.5480	0.4760	108.2533
1986-2005	-46.1846	-1.0003	-0.3937	0.6226	68.8601
1987-2006	-174.8356	-1.0003	-0.3637	0.6152	71.3003
1988-2007	-149.5205	-1.0000	-0.3103	0.6363	68.6620
1989-2008	-612.8425	-1.0000	-0.3210	0.6858	2892.6317
1990-2009	-223.8490	-1.0000	-0.4303	0.5852	99.5186
1991-2010	-419.4585	-1.0000	-0.4543	0.4831	51.6785
$\hat{\sigma}^2$					
1980-1999	0.0000	0.0004	0.0010	0.0021	0.0831
1981-2000	0.0000	0.0004	0.0009	0.0021	0.0570
1982-2001	0.0000	0.0004	0.0010	0.0020	0.0398
1983-2002	0.0000	0.0004	0.0009	0.0019	0.0383
1984-2003	0.0000	0.0004	0.0009	0.0018	0.0360
1985-2004	0.0000	0.0003	0.0008	0.0017	0.0523
1986-2005	0.0000	0.0003	0.0008	0.0017	0.0442
1987-2006	0.0000	0.0003	0.0008	0.0016	0.0440
1988-2007	0.0000	0.0003	0.0008	0.0017	0.0420
1989-2008	0.0000	0.0003	0.0008	0.0017	0.1113
1990-2009	0.0000	0.0003	0.0008	0.0018	0.1735
1991-2010	0.0000	0.0003	0.0008	0.0018	0.1686

Note: Entries equal to 0.0000 imply that the corresponding values are smaller than  $5 \times 10^{-5}$ .

### D.3 RESTRICTED MODELS

Table 29: Comparison of restricted models: number of winnings in the test for SPA.

h	MSE			MAE		
	Benchmark			Benchmark		
	AMIK M2	OU M2	ARMA M2	AMIK M2	OU M2	ARMA M2
1	<b>436</b>	<b>436</b>	252	419	<b>440</b>	241
2	<b>405</b>	387	256	<b>414</b>	385	253
3	<b>389</b>	347	260	<b>398</b>	371	267
	AMIK M3	OU M3	ARMA M3	AMIK M3	OU M3	ARMA M3
1	423	<b>436</b>	-	426	<b>429</b>	-
2	<b>423</b>	398	-	<b>424</b>	404	-
3	<b>421</b>	372	-	<b>426</b>	373	-
	AMIK M4	OU M4	ARMA M4	AMIK M4	OU M4	ARMA M4
1	389	<b>443</b>	-	388	<b>454</b>	-
2	384	<b>395</b>	-	393	<b>413</b>	-
3	<b>389</b>	368	-	<b>385</b>	383	-

*Note:* The null hypothesis is that the benchmark model is not outperformed by other models in terms of MSE and MAE, respectively. Entries refer to the number of firms for which the null hypothesis cannot be rejected on the 10 percent level. For each loss function the highest number of winnings is shown in boldface.



## BIBLIOGRAPHY

---

- Acemoglu, Daron, Vasco Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi (2012). „The network origins of aggregate fluctuations.“ *Econometrica* 80.5, pp. 1977–2016.
- Al-Malkawi, Husam-Aldin Nizar, Michael Rafferty, and Rekha Pillai (2010). „Dividend policy: a review of theories and empirical evidence.“ *International Bulletin of Business* 9, pp. 171–200.
- Albrecht, Steve, Larry Lookabill, and James McKeown (1977). „The time series properties of annual earnings.“ *Journal of Accounting Research* 15.2, pp. 226–244.
- Alexander, Sidney (1949). „The effect of size of manufacturing corporation on the distribution of the rate of return.“ *The Review of Economics and Statistics* 31.3, pp. 229–235.
- Alfarano, Simone and Mishael Milaković (2008). „Does classical competition explain the statistical features of firm growth?“ *Economics Letters* 101.3, pp. 272–274.
- Alfarano, Simone, Mishael Milaković, Albrecht Irle, and Jonas Kauschke (2012). „A statistical equilibrium model of competitive firms.“ *Journal of Economic Dynamics and Control* 36.1, pp. 136–149.
- Alfarano, Simone, Einar Erlingsson, Marco Raberto, and Hlynur Stefánsson (2013). „On the distributional properties of size, profit and growth of Icelandic firms.“ *Journal of Economic Interaction and Coordination* 8.1, pp. 57–74.
- Allen, David and Heazry Salim (2005). „Forecasting profitability and earnings: a study of the UK market (1982–2000).“ *Applied Economics* 37.17, pp. 2009–2018.
- Almus, Matthias and Eric Nerlinger (2000). „Testing Gibrat’s law for young firms - empirical results for West Germany.“ *Small Business Economics* 15.1, pp. 1–12.
- Amaral, Luís, Sergey Buldyrev, Shlomo Havlin, Philipp Maass, Michael Salinger, Eugene Stanley, and Michael Stanley (1997). „Scaling behavior in economics: the problem of quantifying company growth.“ *Physica A* 244.1, pp. 1–24.
- Amato, Louis and Ronald Wilder (1985). „The effects of firm size on profit rates in U.S. manufacturing.“ *Southern Economic Journal* 52.1, pp. 181–190.
- Anderson, Theodore and Donald Darling (1952). „Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes.“ *The Annals of Mathematical Statistics* 23.2, pp. 193–212.
- Anscombe, Francis and William Glynn (1983). „Distribution of kurtosis statistic for normal statistics.“ *Biometrika* 70.1, pp. 227–234.
- Aoki, Masanao and Hiroshi Yoshikawa (2007). *Reconstructing macroeconomics: a perspective from statistical physics and combinatorial stochastic processes*. Cambridge, MA: Cambridge University Press.
- Axtell, Robert (2001). „Zipf distribution of U.S. firm sizes.“ *Science* 293.5536, pp. 1818–1820.

- Baginski, Stephen, Kenneth Lorek, Lee Willinger, and Bruce Branson (1999). „The relationship between economic characteristics and alternative annual earnings persistence measures.“ *The Accounting Review* 74.1, pp. 105–120.
- Bain, Joe (1951). „Relation of profit rate to industry concentration: American manufacturing, 1936-1940.“ *Quartely Journal of Economics* 65.3, pp. 293–324.
- (1956). *Barriers to new competition*. Cambridge, MA: Harvard University Press.
- Ball, Ray and Ross Watts (1972). „Some time series properties of accounting income.“ *The Journal of Finance* 27.3, pp. 663–681.
- Barney, Jay (1991). „Firm Resources and sustained competitive advantage.“ *Journal of Management* 17.1, pp. 99–120.
- (2001). „Resource-based theories of competitive advantage: a ten-year retrospective on the resource-based view.“ *Journal of Management* 27.6, pp. 643–650.
- Bauman, Mark (2014). „Forecasting operating profitability with DuPont analysis: further evidence.“ *Review of Accounting and Finance* 13.2, pp. 191–205.
- Baumol, William (1967). *Business behavior, value and growth*. New York, NY: Harcourt, Brace & World.
- Bickel, David (2002). „Robust estimators of the mode and skewness of continuous data.“ *Computational Statistics and Data Analysis* 39.2, pp. 153–163.
- Blanchard, Olivier, Changyong Rhee, and Lawrence Summers (1993). „The stock market, profit, and investment.“ *Quartely Journal of Economics* 108.1, pp. 115–136.
- Boeri, Tito and Ulrich Cramer (1992). „Employment growth, incumbents and entrants.“ *International Journal of Industrial Organization* 10.4, pp. 545–565.
- Bottazzi, Giulio and Angelo Secchi (2003). „Common properties and sectoral specificities in the dynamics of U.S. manufacturing companies.“ *Review of Industrial Organization* 23.3, pp. 217–232.
- (2005). „Growth and diversification patterns of the worldwide pharmaceutical industry.“ *Review of Industrial Organization* 26.2, pp. 195–216.
- (2006). „Explaining the distribution of firm growth rates.“ *The RAND Journal of Economics* 37.2, pp. 235–256.
- Bottazzi, Giulio, Giovanni Dosi, Marco Lippi, Fabio Pammolli, and Massimo Riccaboni (2001). „Innovation and corporate growth in the evolution of the drug industry.“ *International Journal of Industrial Organization* 19.7, pp. 1161–1187.
- Bottazzi, Giulio, Elena Cefis, and Giovanni Dosi (2002). „Corporate growth and industrial structures: some evidence from the Italian manufacturing industry.“ *Industrial and Corporate Change* 11.4, pp. 705–723.
- Bottazzi, Giulio, Angelo Secchi, and Federico Tamagni (2008). „Productivity, profitability and financial performance.“ *Industrial and Corporate Change* 17.4, pp. 711–751.
- Bottazzi, Giulio, Alexander Coad, Nadia Jacoby, and Angelo Secchi (2011). „Corporate growth and industrial dynamics: evidence from French manufacturing.“ *Applied Economics* 43.1, pp. 103–116.
- Box, George and Gwilym Jenkins (1970). *Time series analysis - forecasting and control*. San Francisco: Holden Day.

- Box, George and David Pierce (1970). „Distribution of residual autocorrelations in autoregressive-integrated moving average time series models.“ *Journal of the American Statistical Association* 65.332, pp. 1509–1526.
- Brown, Lawrence and Michael Rozeff (1979). „Univariate time-series models of quarterly accounting earnings per share: a proposed model.“ *Journal of Accounting Research* 17.1, pp. 179–189.
- Burgstahler, David and Ilia Dichev (1997). „Earnings management to avoid earnings decreases and losses.“ *Journal of Accounting and Economics* 24.1, pp. 99–126.
- Cable, John and Dennis Mueller (2008). „Testing for persistence of profits' differences across firms.“ *International Journal of the Economics of Business* 15.2, pp. 201–228.
- Callen, Jeffrey, Sherman Cheung, Clarence Kwan, and Patrick Yip (1993). „An empirical investigation of the random character of annual earnings.“ *Journal of Accounting, Auditing & Finance* 8.2, pp. 151–162.
- Campbell, John, Andrew Lo, and Craig MacKinlay (1996). *The econometrics of financial markets*. Princeton, NJ: Princeton University Press.
- Capon, Noel, John Farley, and Scott Hoenig (1990). „Determinants of financial performance: a meta-analysis.“ *Management Science* 36.10, pp. 1143–1159.
- Carvalho, Vasco and Xavier Gabaix (2013). „The great diversification and its undoing.“ *American Economic Review* 103.5, pp. 1697–1727.
- Chesher, Andrew (1979). „Testing the law of proportionate effect.“ *The Journal of Industrial Economics* 27.4, pp. 403–411.
- Cho, Hee-Hae and Vladimir Pucik (2005). „Relationship between innovativeness, quality, growth, profitability, and market value.“ *Strategic Management Journal* 26.6, pp. 555–575.
- Chung, Kee and Stephen Pruitt (1994). „A simple approximation of Tobin's q.“ *Financial Management* 23.3, pp. 70–74.
- Ciccarone, Giuseppe (2004). „Finance and the Cambridge equation.“ *Review of Political Economy* 16.2, pp. 163–177.
- Coad, Alex, Agusti Segarra, and Mercedes Teruel (2013). „Like milk or wine: does firm performance improve with age?“ *Structural Change and Economic Dynamics* 24.C, pp. 173–189.
- Coad, Alexander (2007). „A closer look at serial growth rate correlation.“ *Review of Industrial Organization* 31.1, pp. 69–82.
- Cochrane, John (1991). „A critique of the application of unit root tests.“ *Journal of Economic Dynamics and Control* 15.2, pp. 275–284.
- Collins, William and William Hopwood (1980). „A multivariate analysis of annual earnings forecasts generated from quarterly forecasts of financial analysts and univariate time series models.“ *Journal of Accounting Research* 18.2, pp. 390–406.
- Cont, Rama (2001). „Empirical properties of asset returns: stylized facts and statistical issues.“ *Quantitative Finance* 1.2, pp. 223–236.
- Cox, John, Jonathan Ingersoll, and Stephen Ross (1985). „A theory of the term structure of interest rates.“ *Econometrica* 53.2, pp. 385–407.
- Cramér, Harald (1928). „On the composition of elementary errors.“ *Scandinavian Actuarial Journal* 1928.1, pp. 13–74, 141–180.

- Cuaresma, Jesus and Adelina Gschwandtner (2006). „The competitive environment hypothesis revisited: non-linearity, nonstationarity and profit persistence.“ *Applied Economics* 38.4, pp. 465–472.
- Danis, András, Daniel Retzl, and Toni Whited (2014). „Refinancing, profitability, and capital structure.“ *Journal of Financial Economics* 114.3, pp. 424–443.
- Deloof, Marc (2003). „Does working capital management affect profitability of Belgian firms?“ *Journal of Business Finance & Accounting* 30.3, pp. 573–587.
- Demsetz, Harold (1973). „Industry structure, market rivalry, and public policy.“ *The Journal of Law and Economics* 16.1, pp. 1–9.
- Diebold, Francis (2006). *Elements of forecasting*. Mason, Ohio: South Western.
- Diebold, Francis and Roberto Mariano (1995). „Comparing predictive accuracy.“ *Journal of Business & Economic Statistics* 13.3, pp. 253–263.
- Dosi, Giovanni, Sebastien Lechevalier, and Angelo Secchi (2000). „Introduction: inter-firm heterogeneity - nature, sources and consequences for industrial dynamics.“ *Industrial and Corporate Change* 19.6, pp. 1867–1890.
- Eljelly, Abuzar (2004). „Liquidity-profitability tradeoff: an empirical investigation in an emerging market.“ *International Journal of Commerce and Management* 14.2, pp. 48–61.
- Erickson, Timothy and Toni Whited (2000). „Measurement error and the relationship between investment and “q”.“ *Journal of Political Economy* 108.5, pp. 1027–1357.
- Eubank, Randall (1999). *Spline smoothing and nonparametric regression*. New York, NY: Dekker.
- Fairfield, Patricia and Teri Yohn (2001). „Using asset turnover and profit margin to forecast changes in profitability.“ *Review of Accounting Studies* 6.4, pp. 371–385.
- Fairfield, Patricia, Richard Sweeney, and Teri Yohn (1996). „Accounting classification and the predictive content of earnings.“ *The Accounting Review* 71.3, 337–355s.
- Fairfield, Patricia, Sundaresh Ramnath, and Teri Yohn (2009). „Do industry-level analyses improve forecasts of financial performance?“ *Journal of Accounting Research* 47.1, pp. 147–176.
- Fama, Eugene (1970). „Efficient capital markets: A review of theory and empirical work.“ *The Journal of Finance* 25.2, pp. 383–417.
- (1991). „Efficient capital markets: II.“ *The Journal of Finance* 46.5, pp. 1575–1617.
- Fama, Eugene and Kenneth French (2000). „Forecasting profitability and earnings.“ *Journal of Business* 73.2, pp. 161–178.
- (2002). „Testing trade-off and pecking order predictions about dividends and debt.“ *Review of Financial Studies* 15.1, pp. 1–33.
- Fama, Eugene and James MacBeth (1973). „Risk, return, and equilibrium: empirical tests.“ *Journal of Political Economy* 81.3, pp. 607–636.
- Farjoun, Emmanuel and Moshé Machover (1983). *Laws of chaos: a probabilistic approach to political economy*. London: Verso.
- Feltham, Gerald and James Ohlson (1995). „Valuation and clean surplus accounting for operating and financial activities.“ *Contemporary Accounting Research* 11.2, pp. 689–731.



- Foley, Duncan (1994). „A statistical equilibrium theory of markets.“ *Journal of Economic Theory* 62.2, pp. 321–345.
- (2006). *Adam's Fallacy*. Cambridge, MA: Harvard University Press.
- Foster, George (1977). „Quarterly accounting data: time-series properties and predictive-ability results.“ *The Accounting Review* 52.1, pp. 1–21.
- Freeman, Robert, James Ohlson, and Stephen Penman (1982). „Book rate-of-return and prediction of earnings changes: an empirical investigation.“ *Journal of Accounting Research* 20.2, pp. 639–653.
- Fuller, Wayne (1996). *Introduction to statistical time series*. New York, NY: Wiley.
- Gabaix, Xavier (2011). „The granular origins of aggregate fluctuations.“ *Econometrica* 79.3, pp. 733–772.
- García-Teruel, Pedro and Pedro Martínez-Solano (2007). „Effects of working capital management on SME profitability.“ *International Journal of Managerial Finance* 3.2, pp. 164–177.
- Gardiner, Crispin (2009). *Stochastic methods - a handbook for the natural and social sciences*. Berlin: Springer.
- Garibaldi, Ubaldo and Enrico Scalas (2010). *Finitary probabilistic methods in econophysics*. Cambridge, UK: Cambridge University Press.
- Geroski, Paul and Alexis Jacquemin (1988). „The persistence of profits: a European comparison.“ *The Economic Journal* 98.391, pp. 375–389.
- Geroski, Paul and Mariana Mazzucato (2002). „Learning and the sources of corporate growth.“ *Industrial and Corporate Change* 11.4, pp. 623–644.
- Geroski, Paul, Stephen Machin, and Christopher Walters (1997). „Corporate growth and profitability.“ *The Journal of Industrial Economics* 45.2, pp. 171–189.
- Ghongadze, Jaba and Thomas Lux (2012). „Modelling the dynamics of EU economic sentiment indicators: an interaction-based approach.“ *Applied Economics* 44.24, pp. 3065–3088.
- Goddard, John and John Wilson (1999). „The persistence of profit: a new empirical interpretation.“ *International Journal of Industrial Organization* 17.5, pp. 663–687.
- Goddard, John, John Wilson, and Peter Blandon (2002). „Panel tests of Gibrat's law for Japanese manufacturing.“ *International Journal of Industrial Organization* 20.3, pp. 415–433.
- Goddard, John, Manouche Tavakoli, and John Wilson (2005). „Determinants of profitability in European manufacturing and services: evidence from a dynamic panel model.“ *Applied Financial Economics* 15.18, pp. 1269–1282.
- Goldstein, Itay and Liyan Yang (2014). „Market efficiency and real efficiency: the connect and disconnect via feedback effects.“ Rotman School of Management Working Paper No. 2378120.
- Griffin, Paul (1977). „The time-series behavior of quarterly earnings: preliminary evidence.“ *Journal of Accounting Research* 15.1, pp. 71–83.
- Gschwandtner, Adelina (2005). „Profit persistence in the very long run: evidence from survivors and excitors.“ *Applied Economics* 37.7, pp. 793–806.
- Hall, Marshall and Leonard Weiss (1967). „Firm size and profitability.“ *The Review of Economics and Statistics* 49.3, pp. 319–331.

- Hamilton, James (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.
- Hanel, Petr and Alain St-Pierre (2002). „Effects of R&D spillovers on the profitability of firms.“ *Review of Industrial Organization* 20.4, pp. 305–322.
- Hansen, Peter (2005). „A test for superior predictive ability.“ *Journal of Business & Economic Statistics* 23.4, pp. 365–380.
- Härdle, Wolfgang, Marlene Müller, Stefan Sperlich, and Axel Werwatz (2004). *Non-parametric and semiparametric models*. Berlin: Springer.
- Hart, Peter and Eleanor Morgan (1977). „Market structure and economic performance in the United Kingdom.“ *The Journal of Industrial Economics* 25.3, pp. 177–193.
- Hayashi, Fumio (1982). „Tobin’s marginal q and average q: a neoclassical interpretation.“ *Econometrica* 50.1, pp. 213–224.
- Hayfield, Tristen and Jeffrey Racine (2008). „Nonparametric econometrics: the np package.“ *Journal of Statistical Software* 27.5.
- Hecht, Jason (2008). „Modelling cross-sectional profitability and capital intensity using panel corrected significance tests.“ *Applied Financial Economics* 18.18, pp. 1501–1513.
- Henderson, Daniel and Christopher Parmeter (2015). *Applied nonparametric econometrics*. New York, NY: Cambridge University Press.
- Horowitz, Ira (1970). „Employment concentration in the common market: an entropy approach.“ *Journal of the Royal Statistical Society* 133.3, pp. 463–479.
- Ito, Keiko and Kyoji Fukao (2010). „Determinants of the profitability of Japanese manufacturing affiliates in China and other regions: does localisation of procurement, sales and management matter?“ *The World Economy* 33.12, pp. 1639–1671.
- Jacquemin, Alexis and Charles Berry (1979). „Entropy measure of diversification and corporate growth.“ *The Journal of Industrial Economics* 27.4, pp. 359–369.
- Johnson, Norman, Samuel Kotz, and Narayanaswamy Balakrishnan (1995). *Continuous univariate distribution: 2*. New York, NY: Wiley.
- Jovanovic, Boyan (1982). „Selection and the evolution of industry.“ *Econometrica* 50.3, pp. 649–670.
- Kaldor, Nicholas (1956). „Alternative theories of distribution.“ *The Review of Economic Studies* 23.2, pp. 83–100.
- Kampen, Nicolaas van (1992). *Stochastic processes in physics and chemistry*. Amsterdam: North-Holland.
- Kotz, Samuel, Tomasz Kozubowski, and Krysztof Podgórski (2001). *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. New York, NY: Birkhäuser.
- Kuiper, Nicolaas (1960). „Tests concerning random points on a circle.“ *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen* A.63, pp. 38–47.
- LeRoy, Stephen and Richard Porter (1981). „The present-value relation: tests based on implied variance bounds.“ *Econometrica* 49.3, pp. 555–574.
- Lev, Baruch (1983). „Some economic determinants of time-series properties of earnings.“ *Journal of Accounting and Economics* 5.1, pp. 31–48.

- Levinthal, Daniel (1995). „Strategic management and the exploration of diversity.“ In: *Resource-based and evolutionary theories of the firm: towards a synthesis*. Ed. by Cynthia A. Montgomery. Boston, MA: Kluwer Academic Publishers, pp. 19–42.
- Lintner, John and Robert Glauber (1978). „Higgledy-piggledy growth in America.“ In: *Modern developments in investment management*. Ed. by James Lorie. Hinsdale, IL: Thomson Learning, pp. 594–611.
- Little, Ian (1962). „Higgledy-piggledy growth.“ *Bulletin of the Oxford University Institute of Economics and Statistics* 24.4, pp. 387–412.
- Livan, Giacomo, Simone Alfarano, Mishael Milaković, and Enrico Scalas (2014). „A spectral perspective on excess volatility.“ *Applied Economics Letters* 22.9, pp. 745–750.
- Ljung, Greta and George Box (1978). „On a measure of lack of fit in time series models.“ *Biometrika* 65.2, pp. 297–303.
- Lookabill, Larry (1976). „Some additional evidence on the time series properties of accounting earnings.“ *The Accounting Review* 51.4, pp. 724–738.
- Lotti, Francesca, Enrico Santarelli, and Marco Vivarelli (2003). „Does Gibrat’s law hold among young, small firms?“ *Journal of Evolutionary Economics* 13.3, pp. 213–235.
- Lucas, Robert (1977). „Understanding business cycles.“ *Carnegie-Rochester Conference Series on Public Policy* 5.1, pp. 7–29.
- Lux, Thomas (2009). „Rational forecasts or social opinion dynamics? Identification of interaction effects in a business climate survey.“ *Journal of Economic Behavior and Organization* 72.2, pp. 638–655.
- Lux, Thomas, Eleni Samanidou, Elmar Zschischang, and Dietrich Stauffer (2007). „Agent-based models of financial markets.“ *Reports on Progress in Physics* 70.3, pp. 409–450.
- Majumdar, Sumit (1997). „The impact of size and age on firm-level performance: some evidence from India.“ *Review of Industrial Organization* 12.2, pp. 231–241.
- Marcus, Matityahu (1969). „Profitability and size of firm: some further evidence.“ *The Review of Economics and Statistics* 51.1, pp. 104–107.
- Markides, Constantinos and Peter Williamson (1994). „Related diversification, core competencies and corporate performance.“ *Strategic Management Journal* 15.Special issue: Strategy: Search for new paradigms, pp. 149–165.
- Mason, Edward (1939). „Price and production policies of large scale enterprise.“ *American Economic Review* 29.1, pp. 61–74.
- (1949). „The current state of the monopoly problem in the United States.“ *Harvard Law Review* 62.8, pp. 1265–1285.
- Massey, Frank (1951). „The Kolmogorov-Smirnov test for goodness of fit.“ *Journal of the American Statistical Association* 46.253, pp. 68–78.
- Meese, Richard and Kenneth Rogoff (1983a). „Empirical exchange rate models of the seventies: do they fit out of sample?“ *Journal of International Economics* 14, pp. 3–24.
- (1983b). „The out-of-sample failure of empirical exchange rate models: sampling error or misspecification?“ In: *Exchange Rates and International Macroeconomics*. Ed. by Jacob A. Frenkel. University of Chicago Press and NBER, pp. 67–112.
- Miller, Merton and Franco Modigliani (1961). „Dividend policy, growth, and the valuation of shares.“ *Journal of Business* 34.4, pp. 411–433.

- Mills, Terence and Raphael Markellos (2008). *The econometric modelling of financial time series*. Cambridge, UK: Cambridge University Press.
- Mises, Richard von (1928). *Wahrscheinlichkeit, Statistik und Wahrheit*. Wien: Springer.
- Montgomery, Cynthia (1985). „Product-market diversification and market power.“ *The Academy of Management Journal* 28.4, pp. 789–798.
- Mueller, Dennis (1977). „The persistence of profits above the norm.“ *Economica* 44.176, pp. 369–380.
- (1986). *Profits in the long run*. Cambridge, MA: Cambridge University Press.
- (1990). „The persistence of profits in the United States.“ In: *The dynamics of company profits*. Ed. by Dennis C. Mueller. Cambridge University Press, pp. 35–58.
- Mundt, Philipp (2017). „The ultimate corporate objective is survival.“ Unpublished manuscript.
- Mundt, Philipp, Niels Förster, Simone Alfarano, and Mishael Milaković (2014). „The real versus the financial economy: a global tale of stability versus volatility.“ *Economics: The Open-Access, Open-Assessment E-Journal* 8.2014-17.
- Mundt, Philipp, Mishael Milaković, and Simone Alfarano (2016). „Gibrat’s law redux: think profitability instead of growth.“ *Industrial and Corporate Change* 25.4, pp. 549–571.
- Nadaraya, Elizbar (1964). „On estimating regression.“ *Theory of Probability and its Applications* 9.1, pp. 141–142.
- Naples, Michele and Nahid Aslanbeigui (1996). „What does determine the profit rate? The neoclassical theories presented in introductory textbooks.“ *Cambridge Journal of Economics* 20.1, pp. 53–71.
- Nissim, Doron and Stephen Penman (2001). „Ratio analysis and equity valuation: from research to practice.“ *Review of Accounting Studies* 6.1, pp. 109–154.
- Ohlson, James (1995). „Earnings, book values, and dividends in equity valuation.“ *Contemporary Accounting Research* 11.2, pp. 661–687.
- Ohlson, James and Beate Juettner-Nauroth (2005). „Expected EPS and EPS growth as determinants of value.“ *Review of Accounting Studies* 10.2, pp. 349–365.
- Palepu, Krishna (1985). „Diversification strategy, profit performance and the entropy measure.“ *Strategic Management Journal* 6.3, pp. 239–255.
- Palich, Leslie, Laura Cardinal, and Chet Miller (2000). „Curvilinearity in the diversification-performance linkage: an examination of over three decades of research.“ *Strategic Management Journal* 21.2, pp. 155–174.
- Pasinetti, Luigi (1962). „Rate of profit and income distribution in relation to the rate of economic growth.“ *The Review of Economic Studies* 29.4, pp. 267–279.
- Pastor, Lubos and Pietro Veronesi (2003). „Stock valuation and learning about profitability.“ *The Journal of Finance* 58.5, pp. 1749–1789.
- Peteraf, Margaret (1993). „The cornerstones of competitive advantage: a resource-based view.“ *Strategic Management Journal* 14.3, pp. 179–191.
- Politis, Dimitris and Joseph Romano (1994). „The stationary bootstrap.“ *Journal of the American Statistical Association* 89.428, pp. 1303–1313.
- Porter, Michael (1980). *Competitive strategy: techniques for analysing industries and competitors*. New York: The Free Press.

- Priestley, Maurice and Tata Subba Rao (1969). „A test for non-stationarity of time series.“ *Journal of the Royal Statistical Society* 31.1, pp. 140–149.
- Qian, Gongming, Lee Li, Ji Li, and Zhengming Qian (2008). „Regional diversification and firm performance.“ *Journal of International Business Studies* 39.2, pp. 197–214.
- Rajan, Raghuram and Luigi Zingales (1995). „What do we know about capital structure? Some evidence from international data.“ *The Journal of Finance* 50.5, pp. 1421–1460.
- Reid, Gavin (1995). „Early life-cycle behaviour of micro-firms in Scotland.“ *Small Business Economics* 7.2, pp. 89–95.
- Risken, Hannes (1996). *The Fokker-Planck equation: methods of solutions and applications*. Berlin: Springer.
- Rosenbusch, Nina, Jan Brinckmann, and Andreas Bausch (2011). „Is innovation always beneficial? A meta-analysis of the relationship between innovation and performance in SMEs.“ *Journal of Business Venturing* 26.4, pp. 441–457.
- Rumelt, Richard (1982). „Diversification strategy and profitability.“ *Strategic Management Journal* 3.4, pp. 359–369.
- Russo, Michael and Paul Fouts (1997). „A resource-based perspective on corporate environmental performance and profitability.“ *The Academy of Management Journal* 40.3, pp. 534–559.
- Samuelson, Paul and Franco Modigliani (1966). „The Pasinetti paradox in neoclassical and more general models.“ *The Review of Economic Studies* 33.4, pp. 269–301.
- Santarelli, Enrico, Luuk Klomp, and Roy Thurik (2006). „Gibrat’s law: an overview of the empirical literature.“ In: *Entrepreneurship, Growth, and Innovation*. Ed. by Enrico Santarelli. Vol. 12. International Studies in Entrepreneurship. New York, NY: Springer, pp. 41–73.
- Scharfenaker, Ellis and Gregor Semieniuk (2016). „A statistical equilibrium approach to the distribution of profit rates.“ *Metroeconomica* in press.
- Schmalensee, Richard (1985). „Do markets differ much?“ *American Economic Review* 75.3, pp. 341–351.
- Schwarz, Gideon (1978). „Estimating the dimension of a model.“ *The Annals of Statistics* 6.2, pp. 461–464.
- Scott, David (2015). *Multivariate density estimation*. Hoboken, NJ: Wiley.
- Shackle, George (1967). *The years of high theory: invention and tradition in economic thought 1926-1939*. Cambridge, UK: Cambridge University Press.
- Shapiro, Samuel and Martin Wilk (1965). „An analysis of variance test for normality (complete samples).“ *Biometrika* 52.3/4, pp. 591–611.
- Shiller, Robert (1979). „The volatility of long-term interest rates and expectations models of the term structure.“ *Journal of Political Economy* 87.6, pp. 1190–1219.
- (1981). „Do stock prices move too much to be justified by subsequent changes in dividends.“ *The American Economic Review* 71.3, pp. 421–436.
- Silverman, Bernard (1986). *Density estimation for statistics and data analysis*. London: Chapman and Hall.
- Slater, Stanley and Eric Olson (2002). „A fresh look at industry and market analysis.“ *Business Horizons* 45.1, pp. 15–22.

- Soliman, Mark (2008). „The use of DuPont analysis by market participants.“ *The Accounting Review* 83.3, pp. 823–853.
- Stanley, Michael, Luís Amaral, Sergey Buldyrev, Shlomo Havlin, Heiko Leschhorn, Philipp Maass, Michael Salinger, and Eugene Stanley (1996). „Scaling behaviour in the growth of companies.“ *Nature* 379.29, pp. 804–806.
- Storn, Rainer and Kenneth Price (1997). „Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces.“ *Journal of Global Optimization* 11.4, pp. 341–359.
- Subbotin, Mikhail (1923). „On the law of frequency of errors.“ *Matematicheskii Sbornik* 31, pp. 296–301.
- Sutton, John (1997). „Gibrat’s legacy.“ *Journal of Economic Literature* 35.1, pp. 40–59.
- Taleb, Nassim (2012). *Antifragile: things that gain from disorder*. New York, NY: Random House.
- Tallman, Stephen and Jiatao Li (1996). „Effects on international diversity and product diversity on the performance of multinational firms.“ *The Academy of Management Journal* 39.1, pp. 179–196.
- Teece, David (1981). „Internal organization and economic performance: an empirical analysis of the profitability of principal firms.“ *Journal of Industrial Economics* 30.2, pp. 173–99.
- Tobin, James (1969). „A general equilibrium approach to monetary theory.“ *Journal of Money, Credit and Banking* 1.1, pp. 15–29.
- Toda, Alexis (2012). „The double power law in income distribution: explanations and evidence.“ *Journal of Economic Behavior and Organization* 84.1, pp. 364–381.
- Touchette, Hugo, Erik van der Straeten, and Wolfram Just (2010). „Brownian motion with dry friction: Fokker-Planck approach.“ *Journal of Physics A: Mathematical and Theoretical* 43.44, pp. 1–21.
- Uhlenbeck, George and Leonard Ornstein (1930). „On the theory of Brownian motion.“ *Physical Review* 36.5, pp. 823–841.
- Urzúa, Carlos (1996). „On the correct use of omnibus tests for normality.“ *Economics Letters* 53.3, pp. 247–251.
- Wagner, Friedrich, Mishael Milaković, and Simone Alfarano (2010). „Firm profitability and the network of organization capabilities.“ *Physica A* 389.21, pp. 4769–4775.
- Waring, Geoffrey (1996). „Industry differences in the persistence of firm-specific returns.“ *American Economic Review* 86.5, pp. 1253–65.
- Watson, Geoffrey (1961). „Goodness-of-fit tests on a circle.“ *Biometrika* 48.1/2, pp. 109–114.
- (1964). „Smooth regression analysis.“ *The Indian Journal of Statistics* 26.4, pp. 359–372.
- Watts, Ross and Richard Leftwich (1977). „The time series of annual accounting earnings.“ *Journal of Accounting Research* 15.2, pp. 253–271.
- Weiss, Christoph (1998). „Size, growth, and survival in the upper Austrian farm sector.“ *Small Business Economics* 10.4, pp. 305–312.
- White, Halbert (2000). „A reality check for data snooping.“ *Econometrica* 68.5, pp. 1097–1126.

- Whittington, Geoffrey (1980). „The profitability and size of United Kingdom companies, 1960-74.“ *The Journal of Industrial Economics* 28.4, pp. 335–352.
- Wilks, Samuel (1938). „The large-sample distribution of the likelihood ratio for testing composite hypotheses.“ *The Annals of Mathematical Statistics* 9.1, pp. 60–62.