

DISS. ETH NO. 25002

---

# Constrained Stochastic Processes in Complex Socio-Economic Systems

---

A thesis submitted to attain the degree of  
DOCTOR OF SCIENCES of ETH ZURICH  
(Dr. sc. ETH Zurich)

*presented by:*

Sandro Claudio Lera

MSc ETH Physics

born September 11, 1990

citizen of Thalwil (ZH), Switzerland

---

*accepted on the recommendation of:*

*Examiner:* Prof. Dr. Didier Sornette - ETH Zurich

*Co-Examiner:* Prof. Dr. Stefano Battiston - University of Zurich



# Acknowledgements

Many people have accompanied me the past three years, contributing in one way or another to making my time as a doctoral student such a beautiful and memorable journey. First and foremost, I would like to thank my supervisor and mentor Prof. Didier Sornette. It is due to his continuous support, endless inputs at all scales, ideas, and effort that I am able to present this diversified thesis. His attitude and immense knowledge have influenced me positively far beyond academic research. My warmest thanks go to Prof. Misako and Hideki Takayasu, who have invited me to Tokyo and have been more than generous collaborators and hosts. It would not have been nearly as much fun without fellow PhD student, colleague, flatmate, and most importantly friend Dionysios Georgiadis. Our night long discussions led to many insights and interesting developments. It was the work with my friend Dr. Matthias Leiss that made me realize how powerful a scientific collaboration can be, and it was a pleasure learning from his postdoctoral guidance. I am grateful for the generous and efficient support from the Centre for Future Resilient Systems at the Singapore-ETH Centre, in particular by Aaron Ang and Prof. Hans Heinemann. Last but not least, I thank Prof. Stefano Battiston and Prof. Antoine Bommier for participating in my PhD defense, and Judith Holzheimer for her reliable assistance towards the completion of my thesis.

Singapore, February 2018

Sandro Lera



# Abstract

This cumulative thesis gathers three different topics under the common umbrella of being stochastic processes subject to one or several constraints. The bulk of the thesis focuses on target zones, which are fiscal political instruments allowing a central bank to enforce the otherwise free floating exchange rate to not exceed a pre-specified minimum or maximum value. We describe such target zones from three different angles. First, from a purely statistical physics perspective, where we explain how the observed square-root decaying volatility can be rationalized by modeling the exchange rate price as a colloidal Brownian particle embedded in an “order book fluid”. Second, in terms of the classical target zone model by Krugman, which, as we show, captures that same square-root shaped volatility behavior, and furthermore highlights the fact that the imposed price constraint cannot be arbitrated. Third, in terms of option prices. We show how a two-sided target zone is formally equivalent to a portfolio of infinitely compound options. This representation lets us measure the underlying fundamental exchange rate value that would have prevailed in absence of a target zone, and allows us to define a proxy for the perceived target zone credibility in terms of an implied option maturity. We conclude the discussion on target zones by mapping the option representation to the standard model by Krugman.

In the second part of the thesis, we focus on fluctuations of the gross domestic product. Using wavelet transforms, we provide evidence that the distribution of real US GDP per capita growth rate fluctuations is bimodal and heavy-tailed, and we explain how these properties can be seen as the result of a Lévy flight being embedded in a steep confining potential. We interpret the Lévy noise as a representation of firm or sector specific idiosyncratic shocks. The restitution force associated with the steep confining potential well is acting to damp these idiosyncratic shocks, smoothing the effect of individual shocks on the entire economy.

In the third and last part of the thesis, we consider, in a broad sense, branching and colliding random walks. As an example of a colliding random walk, we describe the effect of merger and acquisitions (collisions) on the firm size distribution, that is mostly dominated by the growth of individual firms (random growth). By solving an integro-differential coagulation equation both analytically to first order and numerically, we show that the impact of merger and acquisitions is much less than expected a priori, but in good agreement with empirical observations. In the last chapter of the thesis, we finally present a new type of fitness adjusted preferential attachment model, that includes preferential deletion. We derive an exact condition for a phase transition after which one or a few nodes capture a finite fraction of all links in the infinite networks (sometimes referred to as ‘dragon-kings’). The network topology of the underlying graph can be mapped to a branching stochastic process.



# Zusammenfassung

In dieser kumulativen These werden drei Themengebiete unter dem gemeinsamen Nenner stochastischer Prozesse unter Randbedingungen zusammengefasst. Ein Grossteil der These befasst sich mit sogenannten Zielzonen, also geldpolitische Instrumenten, welche es der Zentralbank erlauben ihre Währung zwischen einem oberen oder unteren Wert zu halten. Wir beschreiben solche Zielzonen aus drei verschiedenen Perspektiven: Zuerst aus Sicht der statistischen Physik, wo wir zeigen, wie die als Quadratwurzel zerfallende Volatilität als Brownsches Teilchens eingebettet in einer "Blotter-Flüssigkeit" gesehen werden kann. Dann innerhalb Krugmans klassischem Zielzonen Modell, welches, wie wir zeigen, mit unserer Vorhersage für die Volatilität übereinstimmt, und auch aufzeigt, dass die Zielzone keine Arbitrage erlaubt. Drittens, mit Hilfe von Optionspreisen. Wir zeigen, dass eine doppelseitige Zielzone formal äquivalent zu einem Portfolio von unendlich ineinander geschachtelter Optionspreise ist. Diese Äquivalenz erlaubt es uns den nicht direkt beobachtbaren Fundamentalwert des Wechselkurses zu bestimmen, welcher in Abwesenheit der Zielzone vorherrschen würde. Wir schliessen das Thema Zielzonen ab indem wir eine mathematische Verallgemeinerung des Krugman Modells zeigen, welches dann mit der Beschreibung des Optionspreismodells übereinstimmt.

In einem zweiten Teil der These betrachten wir Fluktuationen des Bruttoinland Produkts (BIP). Wir zeigen, dass die Verteilungsfunktion der BIP Fluktuationen sowohl bimodal wie auch endlastig sind, was wir als das Resultat eines Lévy-Flugs eingebettet zwischen steilen Potentialwänden modellieren. Wir interpretieren den Lévy Lärm as Representation idiosynkratischer, Firmen-spezifischen Shocks. Die Rücktreibende Kraft assoziiert mit den steilen Potentialwänden dämpft diese einzelnen Shocks und glättet dabei die Auswirkung einzelner Firmen auf die gesamte Wirtschaft.

In einem dritten und letzten Teil befassen wir uns mit verzweigenden und kollidierenden stochastischen Prozessen. Als Beispiel eines kollidierenden Prozesses beschreiben wir den Effekt von Firmenfusionen (Kollisionen) auf die Verteilung von Firmengrössen, welche selbst aufgrund interner Operationen stochastisch wachsen. Indem wir eine Integro-Differentialgleichung sowohl analytisch als auch numerisch lösen, zeigen wir, dass der Einfluss von Firmenfusionen kleiner ist als vielleicht intuitiv erwartet, aber in guter Übereinstimmung mit empirischen Beobachtungen ist. Im letzten Kapitel zeigen wir ein neues Model des Fitness angepassten bevorzugten Anschlusses, welches auch bevorzugtes Löschen einzelner Knoten im Netzwerk beinhaltet. Wir leiten eine exakte Bedingung für einen Phasenübergang her, nach welchem einer oder mehrere Knoten für eine endliche Fraktion aller Verbindungen im unendlichen Netzwerk verantwortlich sind. Die zugrundeliegende Topologie des Netzwerks kann als eine Form eines verzweigenden Zufallsprozesses angesehen werden.





# Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction</b>  | <b>1</b>  |
| <b>2</b> | <b>Currency target zone modeling: An interplay between physics and economics</b>   | <b>7</b>  |
| 2.1      | Target zone arrangements   | 8         |
| 2.2      | A physicist's first approach   | 10        |
| 2.3      | Perturbative solutions of SDEs   | 12        |
| 2.4      | Empirical estimation of drift and volatility   | 24        |
| 2.5      | The no-arbitrage condition   | 27        |
| 2.6      | A physicist's second approach  | 28        |
| 2.7      | Equilibrium in physics and in economics  | 30        |
| 2.8      | Conclusions  | 31        |
| <b>3</b> | <b>Quantitative modelling of the EUR/CHF exchange rate during the target zone regime of September 2011 to January 2015</b> | <b>33</b> |
| 3.1      | Krugman's target zone model  | 35        |
| 3.2      | Previous empirical tests of Krugman's model  | 38        |
| 3.3      | Two new observables to test Krugman's model  | 40        |
| 3.4      | Application to the euro/Swiss franc target zone  | 43        |
| 3.5      | Conclusions  | 53        |
| <b>4</b> | <b>An option representation of currency target zones</b>   | <b>55</b> |
| 4.1      | Target zones as infinite sequences of compound options   | 57        |
| 4.2      | Iterative implementation of mirrored option prices   | 59        |
| 4.3      | Proof of convergence   | 61        |
| 4.4      | Estimating maturity and fundamental value  | 64        |
| 4.5      | Application to the USD/HKD exchange rate   | 66        |
| 4.6      | Conclusions  | 73        |
| <b>5</b> | <b>An explicit mapping of currency target zone models to option prices</b>   | <b>75</b> |
| 5.1      | Phenomenological target zone models  | 76        |
| 5.2      | Target zones from an option perspective  | 78        |
| 5.3      | Relating the two approaches  | 79        |
| 5.4      | Implications of the mapping  | 81        |
| 5.5      | Conclusions  | 83        |
| <b>6</b> | <b>Evidence of a bimodal US GDP growth rate distribution: A wavelet approach</b>   | <b>85</b> |
| 6.1      | Business cycles as Markov autoregressive processes   | 88        |
| 6.2      | Scale free business cycles   | 91        |
| 6.3      | The wavelet transform  | 92        |
| 6.4      | Wavelet analysis of the growth of real US GDP per capita   | 94        |
| 6.5      | Evidence for a robust bimodal structure of distributions of US GDP growth rates  | 96        |
| 6.6      | Conclusions  | 98        |

|           |  |            |
|-----------|--|------------|
| <b>7</b>  | <b>GDP growth rates as confined Lévy flights: towards a unifying macro theory of economic growth rate fluctuations</b> | <b>101</b> |
| 7.1       | Structure of the GDP growth rate distribution . . . . .  | 102        |
| 7.2       | Lévy flights in steep confining potentials . . . . .   | 104        |
| 7.3       | Fitting the Lévy flight parameters . . . . .   | 106        |
| 7.4       | Growth rate fluctuations as confined Lévy flights . . . . .  | 107        |
| 7.5       | Conclusions . . . . .  | 111        |
| <b>8</b>  | <b>Evolution of firm size distributions due to M&amp;A</b>   | <b>113</b> |
| 8.1       | The effect of mergers and acquisitions on Zipf's law . . . . .   | 115        |
| 8.2       | Mergers and acquisitions as a coagulation process . . . . .  | 116        |
| 8.3       | Coagulation equation for heavy tailed distributions . . . . .  | 119        |
| 8.4       | Connecting the coagulation equation to empirical data . . . . .  | 127        |
| 8.5       | Generality of our results . . . . .  | 130        |
| 8.6       | Conclusions . . . . .  | 132        |
| <b>9</b>  | <b>Classification of dragon-king phases in preferential attachment and failure models</b>                              | <b>135</b> |
| 9.1       | Theoretical framework of network growth with fitness-adjusted preferential attachment and deletion . . . . .           | 136        |
| 9.2       | Generic case: The fit get richer . . . . .   | 140        |
| 9.3       | Special case: Dragon-king nodes . . . . .  | 141        |
| 9.4       | Parametric classification of the DK regime . . . . .   | 143        |
| <b>10</b> | <b>Conclusions</b>   | <b>147</b> |

# Chapter 1

## Introduction

Stochasticity, and more specifically the random walk constitutes the pillar of both physics and finance. Both fields have mutually fertilized each other leading to rich mathematical representations of stochastic fluctuations [1, 2]. In physics, fluctuating phenomena form the basis of classical statistical physics, quantum mechanics all the way up to modern string theory [3]. The randomness in market prices is usually argued to be a directed consequence of the no-arbitrage condition [4, 5]. Only recently, more mechanical explanations for the origin of randomness have been provided, in terms of order book dynamics [6, 7], or as the result of the fine-tuned competition between liquidity takers and market makers [8].

This thesis is also concerned with random fluctuations. But instead of pure random walks, we examine the dynamics of random fluctuations in socio-economic systems, subject to different constraints. Each chapter is based on a journal paper [9–16], that has either been already published or is undergoing peer review at the time of writing.

In chapter 2 (based on paper [9]), we examine how the price of an exchange rate, which is a priori a free random walk, is adjusted when the central bank enforces a strict minimum or maximum exchange rate limit (target zone). Concretely, we study the performance of the euro/Swiss-franc exchange rate in the extraordinary period from September 6, 2011, to January 15, 2015, when the Swiss National Bank enforced a minimum exchange rate of 1.20 Swiss francs per euro. We first approach the problem from a physical perspective, but then realize that, in contrast to a purely mechanical description, the price dynamics is more complex than anticipated, as the expected price change is coupled to the price itself. Applying non-parametric estimates, we find that the volatility of the Swiss-franc/euro exchange rate decays like a square root as a function of the distance to the minimum exchange rate of 1.20 Swiss-franc per euro. Using

a recently proposed extended analogy in terms of a colloidal Brownian particle embedded in a fluid of molecules associated with the underlying order book, we derive that, close to the restricting boundary, the same square-root behavior is expected from first-order calculations in hydrodynamics.

Chapter 3 (based on paper [10]) considers target zones from the perspective of economic theories. Specifically, we consider the famous target zone model of Krugman. Despite its simplicity and elegance, empirical evidence for the model has been lacking, not least because it is difficult to capture the predicted non-linear relationship between the observable exchange rate and the non-observable fundamental value. This is why we propose a different approach. By inverting locally the relation between exchange rate and fundamental value, we derive analytical expressions for the conditional volatility and the probability density as a function of the exchange rate. This allows us to examine Krugman's prediction directly from historical data, and, furthermore, enables us to test the smooth pasting condition, which is intimately related to the no-arbitrage condition. We show that the data is well explained by the theory and conclude that Krugman's target zone model holds after all, but apparently only under extreme and sustained pressure that pushes continuously the exchange rate very close to the boundary of the target zone, as was the case for the euro/Swiss-franc target zone.

In chapter 4 (based on paper [11]), we study target zones yet from another angle, in terms of forex options which, by construction, take into consideration the market's expected price change. In presence of a single upper (resp. lower) target boundary, the exchange rate is precisely represented as the sum of a free float and a short (resp. long) position in a call (resp. put) option with strike price at the boundary. To model a target zone with two boundaries, a natural approach consists in describing the exchange rate dynamics as the combination of the two, namely the sum of free float together with a long position in a put written on the lower boundary and a short position in a call option written on the upper boundary, respectively. We show that this first order approximation leads to significant mispricing (as much as 20%) and must be iterated, leading to an infinite sequence of compounded 'mirrored' option prices. We analyze basic properties of such mirrored nested options analytically, describe how to calculate them numerically, and show why it is crucial to take into account higher order corrections in realistic target zones. We argue that this analogy to option prices allows for conceptually simple generalizations that describe different target zone arrangements. We apply our methodology to the estimation of the fundamental value of the Hong Kong dollar that is hidden by the target

zone peg to the US dollar. We also estimate the implied maturity and explain how this parameter serves as direct proxy for target zone credibility.

In chapter 5 (based on paper [12]), we finally present an explicit connection between the phenomenological target zone model from chapter 3 to the option perspective from chapter 4. While the two models are qualitatively similar, they are quantitatively distinct, as we show. We then derive a consistent generalization of the phenomenological model such that it becomes equivalent to the option representation, allowing us to relate a difficult to estimate phenomenological variable in terms of well-defined economic measurables. This neatly combines two so far quantitatively unconnected streams of literature and ends the target zone discussions of this thesis.

The next two chapters are focused on another stochastic time-series: gross-domestic product (GDP) growth rates. Chapter 6 (based on paper [13]) provides novel evidence that the GDP growth rate distribution exhibits a bimodal structure. Using a wavelet transform, we present a quantitative characterisation of the fluctuations of the annualized growth rate of the real US GDP per capita at many scales. The chosen mother wavelet (first derivative of the Gaussian function) applied to the logarithm of the real US GDP per capita provides a robust estimation of the instantaneous growth rate at different scales. We find that business cycles appear at all scales and the distribution of GDP growth rates can be well approximated by a bimodal function associated to a series of switches between regimes of strong growth rate  $\rho_{\text{high}}$  and regimes of low growth rate  $\rho_{\text{low}}$ . The succession of such two regimes compounds to produce a remarkably stable long term average real annualized growth rate of 1.6% from 1800 to 2010 and  $\approx 2.0\%$  since 1950, which is the result of a subtle compensation between the high and low growth regimes that alternate continuously. Thus, the overall growth dynamics of the US economy is punctuated, with phases of strong growth that are intrinsically unsustainable, followed by corrections or consolidation until the next boom starts. We interpret these findings within the theory of “social bubbles” and argue as a consequence that estimations of the cost of the 2008 crisis may be misleading. We also interpret the absence of strong recovery since 2008 as a protracted low growth regime  $\rho_{\text{low}}$  associated with the exceptional nature of the preceding large growth regime.

The results from chapter 6 set the stage for chapter 7 (based on paper [14]), where a new model that combines economic growth rate fluctuations at the microscopic and macroscopic level is presented. At the microscopic level, firms are growing at different rates while also being exposed to idiosyncratic shocks at the firm and sector level. We describe such fluctuations as independent

Lévy-stable fluctuations, varying over multiple orders of magnitude. These fluctuations are aggregated and measured at the macroscopic level in averaged economic output quantities such as GDP. A fundamental question is thereby to what extent individual firm size fluctuations can have a noticeable impact on the overall economy. We argue that this question can be answered by considering the Lévy fluctuations as embedded in a steep confining potential well, ensuring nonlinear mean-reversal behavior, without having to rely on microscopic details of the system. The steepness of the potential well directly controls the extent towards which idiosyncratic shocks to firms and sectors are damped at the level of the economy. Additionally, the theory naturally accounts for business cycles, represented in terms of a bimodal economic output distribution, and thus connects two so far unrelated fields in economics. By analyzing 200 years of US GDP growth rates, we find that the model is in good agreement with the data.

The final two chapters deal with a different type of constraints, that we can summarize as colliding and branching random walks. In chapter 8 (based on paper [15]), we study the influence of merger and acquisitions on the firm size distribution. In other words, we consider the evolution of multiple individual stochastic processes (firm sizes), subject to occasional ‘interaction constraints’ in the form of merger and acquisitions. As such, that chapter can be seen as a special case of ‘colliding random walks’. The distribution of firm sizes is known to be heavy tailed. In order to account for this stylized fact, previous economic models have focused mainly on growth through investments in a company’s own operations (internal random growth). Thereby, the impact of mergers and acquisitions (M&A) on the firm size (external growth, ‘collisions’) is often not taken into consideration, notwithstanding its potential large impact. In the last chapter, we make a first step into accounting for M&A. Specifically, we describe the effect of mergers and acquisitions on the firm size distribution in terms of an integro-differential equation. This equation is subsequently solved both analytically and numerically for various initial conditions, which allows us to account for different observations of previous empirical studies. In particular, it rationalises shortcomings of past work by quantifying that mergers and acquisitions develop a significant influence on the firm size distribution only over time scales much longer than a few decades. This explains why M&A has apparently little impact on the firm size distributions in existing data sets. Our approach is very flexible and can be extended to account for other sources of external growth, thus contributing towards a holistic understanding of the distribution of firm sizes.

In chapter 9 (based on paper [16]), a new case of a preferential attachment model is con-

sidered. The probability that an already existing node in a network acquires a link to a new node is proportional to the product of its intrinsic fitness and its degree. We enrich this already known model by preferential deletion, which removes nodes at random with probability proportional to their fitness to some exponent. Under ‘normal’ conditions, the resulting node degree distribution is an asymptotic power-law (scale-free regime). We derive an exact condition for a phase transition after which one or a few nodes capture a finite fraction of all links in the infinite networks (dragon-king regime). By approximately ‘parametrizing’ the space of fitness distributions through the beta-density, we then show phase-diagrams that separate the two regimes. The topological structure of the generated network can be seen as a special case of a branching random walk. Chapter 10 concludes by collecting all the major insights gained in these previous chapters.





## Chapter 2

# Currency target zone modeling: An interplay between physics and economics †

Perhaps not apparent at first glance, physics and economics have been life-long companions during their mutual development of concepts and methods emerging in both fields. There has been much mutual enrichment and cross-fertilization. Since the beginning of the formulation of the scientific approach in the physical and natural sciences, economists have taken inspiration from physics, in particular in its success in describing natural regularities and processes [2]. Already in 1776, Adam Smith formulated his ‘Inquiry into the Nature and Causes of the Wealth of Nations’ inspired by the ‘Philosophiae Naturalis Principia Mathematica’ (1687) of Sir Isaac Newton, which specifically stresses the notion of causative forces. Reciprocally, physics has been inspired several times by observations in economics. A prominent example of this kind is the theory of Brownian motion and random walks. In order to model the apparent random walk motion of bonds and stock options in the Paris stock market, mathematician Louis Bachelier [4] developed in his thesis the mathematical theory of diffusion. He solved the parabolic diffusion equation five years before Albert Einstein [17] established the theory of Brownian motion based on the same diffusion equation, also underpinning the theory of random walks. These two works have ushered research on mathematical descriptions of fluctuation phenomena in statistical physics, of quantum fluctuation processes in elementary particles-fields physics, on the one hand, and of financial prices on the other hand, both anchored in the random walk model and Wiener process.

---

†This chapter is based on reference [9].

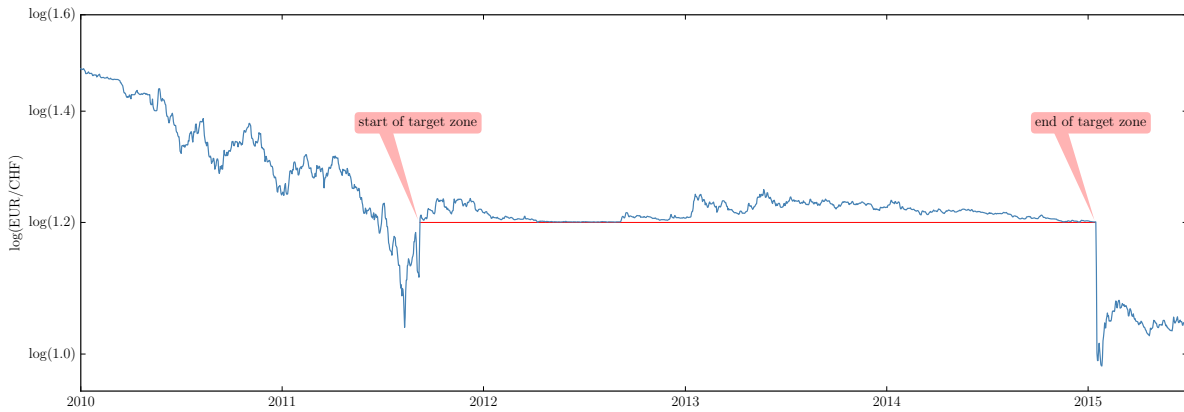


Figure 2.1: Euro/Swiss franc exchange rate between January 1, 2010 and January 14, 2015. On September 6, 2011, the Swiss National Bank (SNB) officially announced its decision to enforce a minimum of 1.20 Swiss francs per euro (red line) by buying euros and selling Swiss francs in unlimited amounts if necessary. The almost immediate appreciation of the Swiss franc back to its pre-target zone value around 1.0 Swiss francs per euro on January 15, 2015, when the SNB announced its abandoning of the target zone policy, indicates how strong the effect of the target zone was.

In this chapter, we extend this analogy even further and describe a restricted Brownian motion both from a physical and an economical perspective, paying close attention to the interplay between the two fields.

## 2.1 Target zone arrangements

On September 6, 2011, the Swiss National Bank (SNB) enforced a minimum exchange rate of 1.20 Swiss francs (CHF) per euro (EUR), in response to the European debt crisis and a continuously weakening euro. With a level around 1.6 CHF at the introduction of the euro in 1999 and a peak above 1.67 CHF on October 2007, the EUR/CHF exchange rate has been floating freely until it dived to the record low of CHF 1.0070 per euro on August 9, 2011. The Swiss National Bank intervened massively leading to a fast rebound of the euro. On September 6, the SNB announced officially that it would defend the minimum exchange rate of CHF 1.20 by all means (buying euros and selling Swiss francs in unlimited amounts as deemed necessary). The SNB held this policy without any exceptions until January 15, 2015, leading to an exchange rate levelling off between 1.20 and 1.24 CHF per euro, exhibiting a dynamics that is spectacularly different from what is observed for a freely floating currency pair, see Figure 2.1. Such a restriction of an exchange rate is known as a target zone arrangement. In the general setup of a target zone, the central bank announces that the exchange rate between the domestic and a foreign currency (or a weighted

basket thereof) as determined by the free market will not be let to exceed some pre-defined upper or lower boundary. As long as the exchange rate is well within the target zone boundaries, the exchange rate is (in the ideal case) allowed to fluctuate freely, controlled only by supply and demand on the foreign-exchange market (FX).<sup>1</sup> The central bank pledges to intervene as soon as the exchange rate approaches closely the lower or upper limit of the band. We follow here the usual convention that the exchange rate denotes the amount of domestic currency that is needed to buy one unit of foreign currency. Hence, we call the lower (upper) limit also the strong (weak) side of the target zone. Once the exchange rate approaches the strong limit from above, the central bank intervenes on the FX by selling its domestic currency to buy foreign exchange, thus increasing supply of the domestic and demand for the foreign currency. This pushes the exchange rate back above its lower limit. The intervention on the weak side is analogous.

A target zone is not a new concept but has been implemented by various central banks in the past decades.<sup>2</sup> Most notably, under the Bretton Woods system, countries would allow to fluctuate their currencies only by 1% around its parity value. To stabilise inflation and diminish large fluctuations between European currencies after the demise of the Bretton Woods system, the European Monetary System (EMS) was established. Members of the European Economic Community (EEC) would restrict fluctuations to  $\pm 2.25\%$  (narrow band) and  $\pm 6\%$  (wide band), respectively. Accordingly, a broad spectrum of papers have aimed at accounting for the different target zone arrangements and at describing speculative attacks.

The most general ansatz to model the dynamics of a financial time series is represented by

$$\frac{ds}{dt} = f(s, t) + g(s, t) \eta(t) \quad (2.1)$$

where  $s$  is the logarithm of the exchange rate, with  $\eta$  Gaussian white noise and  $f, g$  are two functions representing respectively the drift or expected return and the volatility (standard deviation). This ansatz is general since higher order derivatives would imply temporal correlations, thus violating the no-arbitrage condition at the heart of the efficient market hypothesis (EMH) [19–21]. Interestingly, it has been pointed out by several physicists that exactly the use of second order differential equations can have its merits even in finance, see [22–24].

In the remainder of this chapter,  $s$  denotes the log value of one euro measured in units of

---

<sup>1</sup> Interventions inside the target band just to lean against the market movement, in particular without the announcement of an official, well-defined target band is also known as a ‘dirty float’, see for instance [18].

<sup>2</sup> However, we shall see in the next chapter that the EUR/CHF target zone from September 2011 to January 2015 can be considered in some sense special by the extreme nature of its dynamics.

Swiss francs, i.e. the EUR/CHF exchange rate. In order to respect causality for the correct calculation of investments performance, this stochastic equation is understood in the Itô-sense. In the following sections, we will investigate what is the predicted shape of  $f$  and  $g$  from different points of view.

## 2.2 A physicist's first approach

Starting from the structure (2.1), we investigate the nature of the minimal ingredients needed to capture the abnormal dynamics observed in Figure 2.1. The aberrant trajectory of the EUR/CHF log-exchange rate  $s(t)$  is clearly embodied by the visible existence of the barrier at  $s = \underline{s} \equiv \ln(1.20)$  and the tendency for  $s(t)$  to remain very close to it between September 2011 and January 2015. The simplest direct application of the GBM model to this situation is to assume that  $s(t)$  continues to follow a simple random walk but now constrained to remain above the impenetrable cap at  $\underline{s}$ . Since such a situation is not intrinsic to finance, but could just as well correspond to a (one-dimensional) physical Brownian particle that is restricted by a wall at  $\underline{s}$ , we can use this analogy to employ well known mathematical tools from physics. Putting a wall constraint on a random walk is known to induce an effective entropic force acting on the particle, resulting from the reduction of path configurations by reflecting all random walks that would cross the wall [25, 26]. In  $1 + 1$  dimensions (one spatial dimension and one temporal dimension), the corresponding entropic repulsive force can be shown to derive from an effective long-range entropic potential  $V_{\text{ENT}} = C/(s - \underline{s})$ , where  $C > 0$  is a constant [27, 28]. Intuitively, this self-similar long-range potential is associated with the relationship between the average distance to the wall and the long wavelengths of the random walks that are suppressed by the rigid impenetrable barrier.

It is easy to see that the two ingredients ‘Brownian motion’ and ‘restricting wall’ cannot account for the sustained proximity of the particle (exchange rate) to the wall (Figure 2.1). We need to add at least one more ingredient to account for this fact, which, in the economic picture, comes from the strong economic ‘pressure’ on the euro resulting from the European crisis, which led to the introduction of the 1.20-cap in the first place. The simplest assumption is to assume a constant physical pressure that pushes the particle towards the wall, corresponding to the linear potential  $V_{\text{ECO}} = F \cdot (s - \underline{s})$  with a constant  $F > 0$ . Together, this yields the following total potential

$$V \equiv V_{\text{ENT}} + V_{\text{ECO}} = \frac{C}{s - \underline{s}} + F \cdot (s - \underline{s}), \quad (2.2)$$

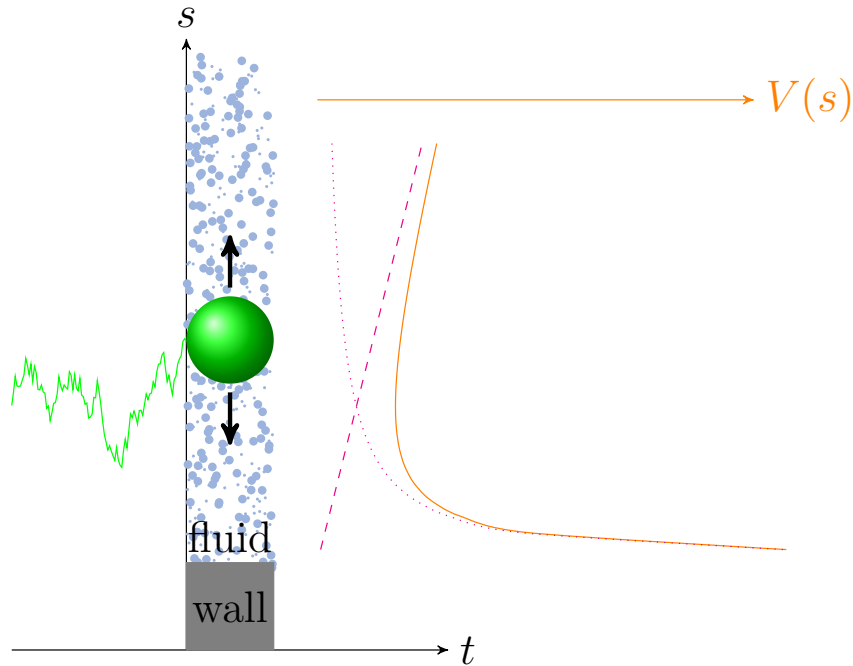


Figure 2.2: Random trajectory (the fluctuating, continuous line on the left) of a one-dimensional Brownian particle moving in a potential  $V(s)$  (continuous line on the right). This potential is the sum of an attractive potential (dashed line) and a repulsive potential (dotted line). From the most simplified physical perspective, one would expect the EUR/CHF exchange rate between September 2011 and January 2015 to be controlled by such force potentials.

depicted in Figure 2.2. The equilibrium position at which expression (2.2) finds its minimum is  $s_{\text{eq}} = \underline{s} + \sqrt{C/\bar{F}}$ : unsurprisingly, the stronger the pressure  $F$  on the euro, the closer is the equilibrium exchange rate to the barrier. Expanding (2.2) around  $s_{\text{eq}}$ , using  $f \equiv -dV/ds$  and inserting this into (2.1) gives to leading orders

$$\frac{ds}{dt} = 3\frac{F^2}{C}(s - s_{\text{eq}})^2 - 2\sqrt{\frac{F^3}{C}}(s - s_{\text{eq}}) + g\eta(t). \quad (2.3)$$

With equation (2.3), we have derived a model aimed at capturing the constrained EUR/CHF dynamics using only a minimal number of ingredients. One way to test the naive hypothesis (2.3) is by calculating the empirical moments from the data and comparing to the theoretical results. Since equation (2.3) cannot be solved analytically, we present in the next section a path integral approach that allows for a perturbative calculation of these moments. The reader only interested in the core results of this chapter can skip the following section and continue with section 2.4.

## 2.3 Perturbative expansion of moments of solutions to stochastic differential equations with path integrals

Above we have encountered a stochastic differential equation (SDE) of the form

$$\dot{X}(t) = -aX(t) + bX(t)^2 + \sqrt{D}\eta(t). \quad (2.4)$$

where  $a, b, D$  are some constants and  $\eta$  is white noise. Finding an analytic solution to (2.4), or, alternatively, its associated Fokker-Planck equation is difficult. Following [29, 30], we discuss a method that enables us to calculate moments and cumulants of solutions to (2.4) based on Feynman path integrals. This approach is unique and conceptually interesting, but rather lengthy. The reader only interested in core results can directly proceed to the next section, where the results are these calculations are presented.

### 2.3.1 Motivation

For a one-dimensional, real random variable  $X$  we define the moment generating function  $Z$  as

$$Z[\lambda] = \langle e^{\lambda X} \rangle = \int dX P(X) e^{\lambda X} \quad (2.5)$$

where  $P$  is the density distribution function of  $X$  and  $\lambda$  a dummy variable. Via

$$\langle X^n \rangle = \frac{1}{Z[0]} \left. \frac{d^n}{d\lambda^n} Z(\lambda) \right|_{\lambda=0} \quad (2.6)$$

we can then calculate the  $n$ -th moment of  $X$  as can be seen immediately by applying the derivative directly on the integrand in (2.5).<sup>3</sup> Note that usually in this context  $Z[0] = 1$  since  $P$  is normalized. However, we must not insist on this condition since this factor cancels out in (2.6) anyway. Furthermore, the cumulant generating function  $Z_C$  is defined as

$$Z_C[\lambda] = \log Z[\lambda]. \quad (2.7)$$

---

<sup>3</sup> Mind the change in notation. In this thesis, we usually denote theoretical expectation values by  $\mathbb{E}$  and empirical mean values by  $\langle \cdot \rangle$ . However, in this appendix we choose to denote theoretical expectation values also by  $\langle \cdot \rangle$  to highlight the formal equivalence between the following calculations and the formalism know from field theories.

By definition, the  $n$ -th cumulant is then given by

$$\langle X^n \rangle_C = \left. \frac{d^n}{d\lambda^n} Z_C(\lambda) \right|_{\lambda=0}. \quad (2.8)$$

Hence, having a closed-form expression for  $Z$  presents a convenient way of calculating moments and cumulants. Let us now consider the special case of a Gaussian distributed variable  $X$  with zero mean and standard deviation  $\sigma$ ,

$$Z[\lambda] = \frac{1}{\sqrt{2\pi}\sigma} \int dX e^{-\frac{X^2}{2\sigma^2} + \lambda X} = Z[0] e^{\lambda a + \frac{\lambda^2 \sigma^2}{2}} \quad (2.9)$$

with  $Z[0] = \sqrt{2\pi}\sigma$ . The second equality follows from completing the square in the exponent. This method generalizes to an  $n$ -dimensional random variable  $\vec{X} = (X_1, \dots, X_n)$  in a natural way through

$$\begin{aligned} Z[\vec{\lambda}] &= \frac{1}{(2\pi)^{n/2} \sqrt{|\det(K)|}} \int \prod_{i=1}^n dX_i \exp \left( -\frac{1}{2} \sum_{j,k} X_j K_{jk}^{-1} X_k + \sum_j \lambda_j X_j \right) \\ &= Z[0] \exp \left( \sum_{j,k} \frac{1}{2} \lambda_j K_{jk} \lambda_k \right) \end{aligned} \quad (2.10)$$

with  $Z[0] = (2\pi \det K)^{n/2}$  and  $K$  the inverse of the covariance matrix. We can now consider the limit to an infinite-dimensional random variable through  $X_i \rightarrow X(t)$  where  $t \in [0, T]$ , leading to the generalization

$$Z[\lambda] \propto \int \mathcal{D}X(t) \exp \left( -\frac{1}{2} \int ds dt X(s) K^{-1}(s, t) X(t) + \int dt \lambda(t) X(t) \right) \quad (2.11)$$

$$= Z[0] \exp \left( \int ds dt \frac{1}{2} \lambda(s) K(s, t) \lambda(t) \right) \quad (2.12)$$

where  $\mathcal{D}X(t) \equiv \lim_{n \rightarrow \infty} \prod_{i=1}^n dx_i$  and  $Z[0] = \lim_{n \rightarrow \infty} (2\pi \det K)^{n/2}$ . In the last step we have again ‘completed the square’. Integrals of type (2.11) which integrate over an entire functional space are commonly called (Feynman) path integrals. Note that  $Z[0]$  is formally infinite but as already argued above, we will see that this divergent pre-factor cancels out in the calculation of moments and cumulants.

### 2.3.2 From Differential Equations to Path Integrals

The above was just a special case of an ‘infinite-dimensional Gaussian distributed stochastic process’. Let us have a look now at how we can generalize this idea in order to apply it to a general SDE of the form

$$\dot{X} = f(X, t) + g(X, t) \eta(t) \quad (2.13)$$

with initial condition  $X_0 = X(t = 0)$ . We rewrite this initial value problem as a probability density

$$P[X(t)|X_0, \eta(t)] = \Delta \left[ \dot{X}(t) - f(X, t) - g(X, t)\eta(t) - X_0\delta(t - t_0) \right] \quad (2.14)$$

where  $\Delta$  is the Dirac delta functional, i.e. the functional generalization of Dirac’s delta distribution  $\delta$ . The definition of the density  $P$  via (2.14) is actually a trivial statement since all we do is writing a probability distribution that is constrained at the solution of the SDE. But this is useful because the  $\delta$ -functional has the path-integral representation

$$\delta(X(t)) = \mathcal{N} \int \mathcal{D}K(t) e^{-\int dt iK(t)X(t)} \equiv \mathcal{N} \int \mathcal{D}\tilde{K}(t) e^{-\int dt \tilde{K}(t)X(t)} \quad (2.15)$$

where we have defined  $\tilde{K}(t) := iK(t)$ . Moreover, we denote by  $\mathcal{N}$  some pre-factor in which we will collect all the (possibly diverging) pre-factors that occur in the following calculations. Recalling that  $\eta$  is white noise gives

$$P[X(t)|X_0] = \mathcal{N} \int \mathcal{D}\eta P[X(t)|X_0, \eta] e^{-\eta^2(t)} \quad (2.16a)$$

$$= \mathcal{N} \int \mathcal{D}\eta \mathcal{D}\tilde{K} \exp \left( - \int dt \left[ \tilde{K}(t) \left( \dot{X}(t) - f(X, t) - X_0\delta(t - t_0) \right) + \tilde{K}(t)g(X, t)\eta(t) - \eta^2(t) \right] \right) \quad (2.16b)$$

$$= \mathcal{N} \int \mathcal{D}\tilde{K}(t) \exp \left( - \int dt \left[ \tilde{K}(t) \left( \dot{X}(t) - f(X, t) - X_0\delta(t - t_0) \right) + \frac{1}{2}\tilde{K}^2(t)g^2(X, t) \right] \right) \quad (2.16c)$$

where in the last step we have completed the square and integrated out the  $\eta$ -dependence. We will drop  $\mathcal{N}$  in the following for simplicity and since it would cancel out in final-results anyway.



Now we can write the moment and cumulant generating functionals for the solution of (2.4) as

$$Z[J, \tilde{J}] = \left\langle \exp \left( \int dt \left( \tilde{J}(t)X(t) + J(t)\tilde{K} \right) \right) \right\rangle \quad (2.17a)$$

$$= \int \mathcal{D}X(t) \mathcal{D}\tilde{K}(t) \exp \left( -S[X, \tilde{K}] + \int dt \tilde{J}(t)X(t) + \int dt J(t)\tilde{K}(t) \right) \quad (2.17b)$$

with the action

$$S[X, \tilde{K}] \equiv \int dt \left[ \tilde{K}(t) \left( \dot{X}(t) - f(X(t), t) - X_0\delta(t - t_0) \right) - \frac{1}{2}\tilde{K}^2(t)g^2(X(t), t) \right]. \quad (2.18)$$

The terms proportional to  $\tilde{J}X$  and  $J\tilde{K}$  in (2.17b) are the so-called source terms. Since  $J$  and  $\tilde{J}$  will be set to zero in any final result we have introduced them without loss of generality. Working with functional derivatives, the source terms turn out useful because the moments can be obtained via

$$\left\langle \prod_{i=1}^m \prod_{j=1}^n X(t_i)\tilde{K}(t_j) \right\rangle = \frac{1}{Z[0,0]} \prod_{i=1}^m \prod_{j=1}^n \frac{\delta}{\delta \tilde{J}(t_i)} \frac{\delta}{\delta J(t_j)} Z \Big|_{J=\tilde{J}=0} \quad (2.19)$$

whereas the cumulants obey

$$\left\langle \prod_{i=1}^m \prod_{j=1}^n X(t_i)\tilde{K}(t_j) \right\rangle_C = \frac{1}{Z[0,0]} \prod_{i=1}^m \prod_{j=1}^n \frac{\delta}{\delta \tilde{J}(t_i)} \frac{\delta}{\delta J(t_j)} \log Z \Big|_{J=\tilde{J}=0}. \quad (2.20)$$

Now, let us apply this formalism specifically to (2.4). The expression for the action (2.18) then reads

$$S[X, \tilde{K}] = \int dt \left[ \tilde{K} \left( \dot{X} + aX - bX^2 - X_0\delta(t - t_0) \right) - \frac{D}{2}\tilde{K}^2 \right]. \quad (2.21)$$

We see immediately that this does not correspond to a free action, meaning that  $X$  and  $\tilde{K}$  appear in such a way that we cannot complete the square and perform the integration as before but we will require perturbative methods instead. We thus separate  $S = S_F + S_I$  where  $S_F$  denotes the free part

$$S_F[X, \tilde{K}] = \int dt \tilde{K} \left( \dot{X} + aX \right) = \int dt \tilde{K} \left( \frac{d}{dt} + a \right) X \quad (2.22)$$

and the rest of the action belongs to the interaction part  $S_I$ . Now we only have to determine the inverse (or rather Green's function)  $G$  of the operator  $\left( \frac{d}{dt} + a \right)$ , i.e. solve for

$$\left( \frac{d}{dt} + a \right) G(t, t') = \delta(t - t'). \quad (2.23)$$

The Green's operator is easily derived by solving the above equation in Fourier space, giving

$$G(t, t') = \Theta(t - t')e^{-a(t-t')} \quad (2.24)$$

with  $\Theta$  the Heaviside function. Now that we know  $G$ , we add source terms to  $S_F$ , i.e. define  $S_0 := S_F + \int dt \tilde{J}(t)X(t) + \int dt J(t)\tilde{K}(t)$ . This yields the generating functional of the free action with source terms

$$Z_F[J, \tilde{J}] = Z_0 \exp \left( \int ds dt \tilde{J}(s)G(s, t)J(t) \right) \quad (2.25)$$

where we have completed the square such that all the source terms become independent of  $X$  and  $\tilde{K}$ . This way the path integral becomes nothing but a (diverging) pre-factor  $Z_0 \equiv Z[J=0]$  which cancels out when we calculate moments and cumulants. The generating functional of the entire theory including the interaction part is given as

$$Z[J, \tilde{J}] = \exp \left( S_I \left[ \frac{\delta}{\delta J}, \frac{\delta}{\delta \tilde{J}} \right] \right) Z_F[J, \tilde{J}]. \quad (2.26)$$

By expanding the exponential up to an arbitrary order and using (2.19) and (2.20), respectively, we can calculate any moment or cumulant with arbitrary precision. It is important to note that a perturbative expansion makes only sense if we can make sure that the coupling constants are small. In our case, this means that we must make sure that  $b, D$  and  $X_0$  ( $a$  belongs to the free action) are small. We discuss the validity of this assumption in section 2.3.4. For now, let us assume that this is indeed the case. We have now all we need to calculate the moments and cumulants. We expand the interacting part of the action as

$$\begin{aligned} \exp \left( S_I[X, \tilde{K}] \right) &= 1 - X_0 \tilde{K}(t_0) - b \int dt \tilde{K}(t)X^2(t) - \frac{D}{2} \int dt \tilde{K}^2(t) \\ &\quad + \frac{1}{2} \left( X_0 \tilde{K}(t_0) + b \int dt \tilde{K}(t)X^2(t) + \frac{D}{2} \int dt \tilde{K}^2(t) \right)^2 + \dots, \end{aligned} \quad (2.27)$$

replace  $X$  by  $\delta/\delta J$  and  $\tilde{K}$  by  $\delta/\delta \tilde{J}$ , plug this into (2.26) and then calculate the cumulants via (2.20).

### 2.3.3 Feynman Diagrams

The above technique to calculate cumulants is possible but also rather tedious. There is a much more convenient way to keep track of the necessary terms by introducing Feynman diagrams.

With some practice, one can quickly derive the Feynman diagrams and the corresponding Feynman rules from the action (2.21). There may be some additional subtleties to the rules when the fields involved in the theory have vectorial or tensorial structure as for instance in quantum chromodynamics. But for (2.21) the derivation is straight forward: The propagator  $G$  corresponds to the free part of (2.21) and is graphically denoted by a straight line:

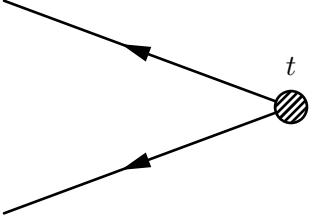
$$\begin{array}{c} t \qquad \qquad \qquad t_0 \\ \longleftarrow \qquad \qquad \qquad \longleftarrow \end{array} = \left\langle \tilde{K}(t)X(t_0) \right\rangle_F = G(t, t_0) \quad (t > t_0). \quad (2.28)$$

The subscript  $F$  means averaging with respect to the free partition function (2.25). We work in the convention that time flows from right to left. The right end of a line then denotes a time  $t_i$  and the left end point denotes a time  $t_f$  with  $t_f > t_i$ . Each summand in the interacting part in (2.21) now constitutes one additional Feynman rule, a so-called vertex. For each  $X$ -field we draw an ingoing and for each  $K$ -field we draw an outgoing line. For the first vertex, we read from (2.21)

$$\begin{array}{c} \diagup \\ \longleftarrow \bullet t \\ \diagdown \end{array} = bX^2(t)\tilde{K}(t). \quad (2.29)$$

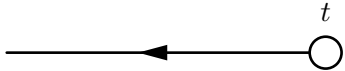
Time  $t$  denotes any time between the time  $t_i$  to the right of the incoming lines and the time  $t_f$  to the left of the outgoing line. Since  $t$  can denote any point with  $t_i < t < t_f$  we will see below that we have to integrate over  $t$ , i.e. consider a superposition of all possibilities  $t \in (t_i, t_f)$ .

The second vertex in (2.21) reads



$$= \frac{D}{2} \tilde{K}^2(t) \quad (2.30)$$

and finally we have a tadpole, i.e. a vertex with no ingoing line,



$$= X_0 \tilde{K}(t) \delta(t - t_0). \quad (2.31)$$

The  $\delta$  arises because this node cannot be connected to the right. Hence, it must denote the initial time  $t_0$ . The usefulness of these diagrams is due to the following statement: Expression (2.19) is equal to the sum of all (topologically distinct) Feynman diagrams with  $m$  ingoing and  $n$  outgoing lines which can be drawn with only the above nodes (2.29)-(2.31). Furthermore, the cumulants (2.20) are given by the sum of all fully connected Feynman diagrams. Here, fully connected means just that all nodes are connected to each other (either directly or via other nodes). This is a powerful statement. It tells us that instead of working with multiple functional derivatives and a myriad of product and chain rules, we can simply draw intuitive diagrams which then correspond to certain mathematical expressions. However, two questions naturally arise: What are distinct diagrams? For a given  $m$  and  $n$ , there seems to be in general an infinite amount of diagrams with  $m$  incoming and  $n$  outgoing lines. Do we have to consider an infinite amount of diagrams? As will become more clear below, there is a one to one correspondence between the number of nodes in a diagram and the order of the expansion in (2.27). We can see from (2.27) that if the coupling constants  $y, b$  and  $D$  are small, only the diagrams with the least amount of nodes will contribute significantly. We will argue in section 2.3.4 that indeed these coupling constants are small and hence we are safe to consider only the diagrams with the smallest amount of nodes.

Regarding the first question, we can look for instance at the node (2.29). We see that there are two possibilities for the two incoming lines to connect to the node. (Line coming from top right going to the top right ‘plug’ of the node or line coming from top right going to the bottom right ‘plug’ of the node. Only the first possibility is depicted in (2.29).) Indeed, these are two topologically distinct diagrams and so they have to be considered as two different diagrams. However, they represent the same mathematical expression. It is therefore easier to just consider each such diagram as one, and then multiply it by its symmetry factor  $\mathcal{S}$ . The symmetry factor accounts exactly for the number of topologically distinct diagrams that represent the same mathematical expression. Another ambiguity arises from the choice  $t$  of the node which must be between  $t_i$  and  $t_f$ . Since each such choice corresponds to a different diagram, we must integrate over all such choices of  $t$ . This leads us finally to the following set of Feynman rules which tell us exactly what mathematical expression we must assign to each diagram:

- (i) Each line between two points of time  $t_i$  and  $t_j$  corresponds to a propagator  $G(t_i, t_j)$  where  $t_i > t_j$ .<sup>4</sup>
- (ii) For  $\mathcal{S}$  distinct ways of connecting edges to vertices that lead to the ‘same’ diagram we account a symmetry factor  $\mathcal{S}$ .
- (iii) For each vertex we multiply by the corresponding pre-factor ( $X_0, b$  or  $D/2$ ) and integrate each vertex over the time-domain  $(t_0, \infty)$ .

It is easy to verify, that these rules yield exactly the same as the approximation of (2.19) and (2.20) via (2.27). Note that for an expansion up to  $k$ -th order in (2.27) we have to consider a factor of  $1/k!$  from the expansion of the exponential. This factor does not show up in our Feynman rules because it cancels exactly with the number of permutations of the vertices that lead to the same diagram.

We are now ready to calculate the cumulants of  $X(t)$  from (2.4). For the variance and skewness of  $X(t)$  we will need to know  $\langle X^n(t) \rangle_C$  for  $n = 2, 3$ . Let us start with  $\langle X^3(t) \rangle_C$ . A first straight forward idea which comes to mind is

---

<sup>4</sup> The case  $t_i = t_j$  (e.g. in loops) automatically drops out since  $\Theta(0) = 0$ , i.e. we work with the left-continuous Heaviside theta function.

$$(2.32)$$

This diagram contributes to  $\langle X^3(t) \rangle$ . However, as argued above, only the fully connected diagrams contribute to the cumulants. (Thinking in terms of (2.20) and (2.27) this can easily be seen as a consequence of the product rule for the functional derivative.) Hence, we can already discard all the diagrams that are not fully connected. For the lowest order contribution to  $\langle X^3(t) \rangle_C$  we then find the following diagram:

$$(2.33)$$

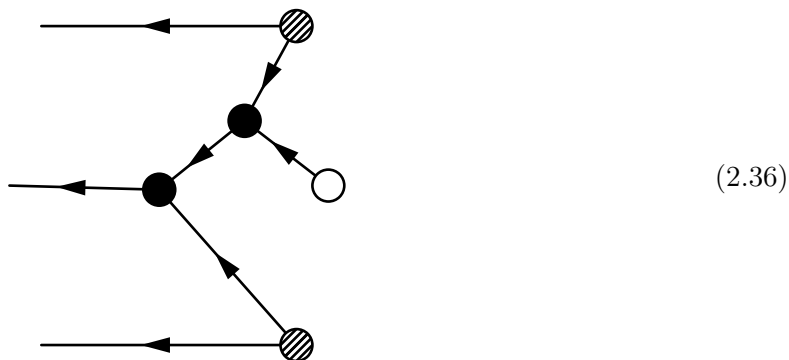
For all the three vertices we have 2 possibilities each to connect the lines to the vertices. This yields a symmetry factor of 8. Hence, we find that

$$\begin{aligned}
 (2.33) &= 8b \frac{D^2}{4} \int dt_1 \int dt_2 \int dt_3 G(t, t_1) G(t, t_2) G(t, t_3) G(t_1, t_2) G(t_1, t_3) \\
 &= \frac{2bD^2}{3a^3} e^{-4at} (e^{at} - e^{at_0})^3 (e^{at} + 3e^{at_0}).
 \end{aligned}
 \tag{2.34}$$

In the limit of large  $t$  we have thus found that the third cumulant is given up to third order by

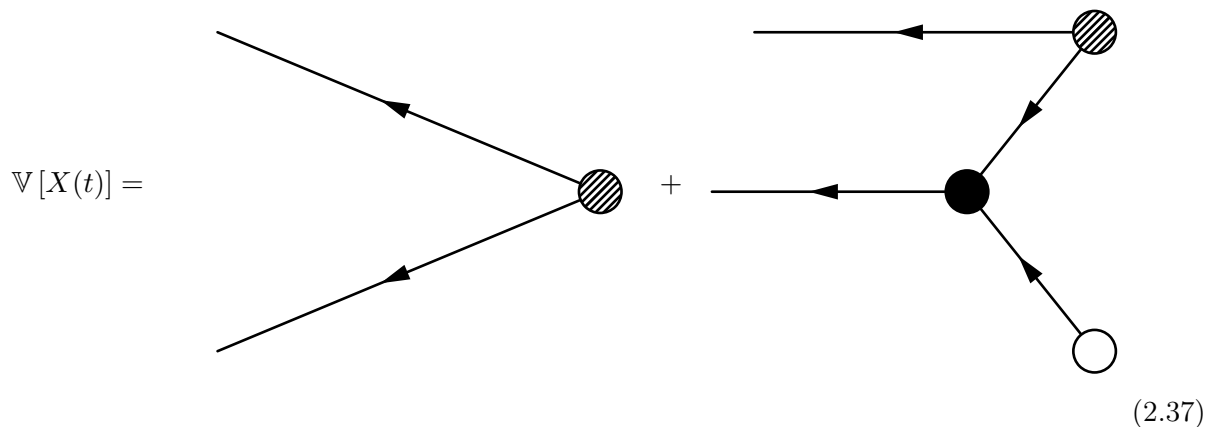
$$\langle X^3(t) \rangle_C = \frac{2bD^2}{3a^3}.
 \tag{2.35}$$

The next higher contribution term is of the form



and so forth. If necessary, we can expand the third cumulant up to arbitrary precision. However, we argue in section 2.3.4 that this is not necessary in our case.

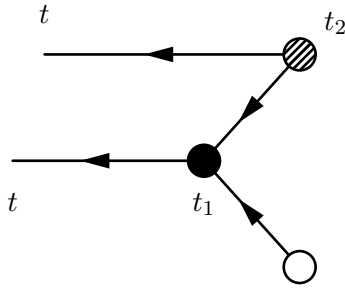
To have everything in the same order we calculate the second cumulant  $\langle X^2(t) \rangle_C$  (which is equal to the variance  $\mathbb{V}[X(t)]$ ) also up to third order:



For the first diagram we calculate

$$\begin{array}{c} t \\ \swarrow \\ \bullet \\ \searrow \\ t \end{array} = D \int dt_1 G(t, t_1)^2 = \frac{D}{2a} (1 - e^{2a(t_0-t)}) \tag{2.38}$$

and for the second one we have



$$= 2byD \int dt_1 \int dt_2 G(t, t_1)G(t, t_2)G(t_1, t_0)G(t_1, t_2)$$

$$= \frac{bX_0D}{a^2} e^{a(t_0-3t)} (e^{at} - e^{at_0})^2. \quad (2.39)$$

Thus, we have found that up to third order

$$\langle X^2(t) \rangle_C = \mathbb{V}(X(t)) = \frac{D}{2a} (1 - e^{2a(t_0-t)}) + \frac{bX_0D}{a^2} e^{a(t_0-3t)} (e^{at} - e^{at_0})^2 \xrightarrow{t \rightarrow \infty} \frac{D}{2a}. \quad (2.40)$$

The skewness  $\gamma$  of a random variable  $X$  is defined as

$$\gamma(X) = \mathbb{E} \left( \left( \frac{X - \mathbb{E}[X]}{\sqrt{\mathbb{V}[X]}} \right)^3 \right) = \frac{\langle X^3(t) \rangle_C}{\mathbb{V}^{3/2}}, \quad (2.41)$$

thus

$$\gamma(X(t)) = \frac{\sqrt{2} b\sqrt{D}}{3 a^{3/2}} \frac{e^{-4at} (e^{at} - e^{at_0})^3 (e^{at} + 3e^{at_0})}{\left(1 - e^{2a(t_0-t)} + \frac{2by}{a} e^{a(t_0-t)} [e^{a(t_0-t)} - 1]^2\right)^{3/2}}. \quad (2.42)$$

In the limit of  $t \rightarrow \infty$  the time dependence cancels out and we are left with

$$\gamma(X(t)) = \frac{\sqrt{2} b\sqrt{D}}{3 a^{3/2}}. \quad (2.43)$$

### 2.3.4 Higher Order Expansions

The method presented in the previous section delivers perturbative results in  $b$ ,  $X_0$  and  $D$  (however, not in  $a$  which is part of the free action) for moments and cumulants of  $X$ . In order to work with the first order result (2.33) for the skewness we have to argue that it is justified to neglect higher order diagrams. This means that we have to argue that the ‘coupling constant’  $b$ ,  $X_0$  and  $D$  are indeed small. Let us consider  $D$  first. Analyzing the time series of the stationary region (Figure 2.1) we see that the fluctuations are of order  $10^{-3}$ . Since the random fluctuations scale with  $\sqrt{D}\eta$  and  $\eta$  is normally distributed with standard deviation equal to unity we see that we must choose  $g = D$  of order  $10^{-6}$ . Hence, perturbation in  $D$  is clearly justified. It is only



reasonable to choose an initial value of similar order such that  $X_0$  must also be of magnitude  $10^{-3}$  or smaller. So perturbation in  $X_0$  is justified. Finally, we have that  $b = 6F^2/c$ . We have shown in section 2.2 that  $F \sim 1/s_{\text{eq}}^2$  and since  $s_{\text{eq}} \gtrsim \log(1.20)$  we see that  $F^2$  is of order  $1/2$ . For  $b$  to be small, we must show that  $c$  is small. This is problematic, since  $c$  is a parameter from the model that we cannot influence or estimate a priori. However, we will show now that perturbation is still justified, as long as  $c$  is reasonably ‘small’.

The third cumulant  $\langle X^3(t) \rangle_C$  is given in first approximation by (2.33). This means first order in  $b$ , second order in  $D$  and zeroth order in  $X_0$ . Since  $D \ll X_0$ , the order of  $D$  is the one that must be kept the lowest since one order higher in  $D$  will already make the diagram several orders of magnitude less important. We can keep  $D$  in second order and expand up to arbitrary order in  $b$  and  $X_0$  according to:

$$+ \dots \quad (2.44)$$

Expanding in this manner we see that if we expand up to  $n$ -th order in  $b$  we are at  $(n - 1)$ -th order in  $X_0$ . However, we know that  $X_0$  is of order  $10^{-3}$  and so as long as  $c$  is of order  $10^2$  or smaller perturbation is still justified. We can also expand in higher orders of  $D$ . In third order of  $D$  we have for instance

$$(2.45)$$

and so forth. However, since  $D$  is of order  $10^{-6}$  we can safely neglect higher order expansions in  $D$ . It is therefore completely justified to consider only contributions from (2.33) to the calculation

of  $\langle X^3(t) \rangle_C$  as long as the value of  $b$  is of order 100 or lower because then multiplication with  $X_0$  shows that each higher order is still one order of magnitude less significant. We could even set  $X_0 = 0$  (and lose some generality) and then  $c$  could take values of order  $10^5$  and the expansions would still be justified. By nature of the problem, we can assume that  $b$  will not exceed such large values since  $b \sim 1/c$  and  $c$  is a constant in the potential  $V = \frac{c}{r} + Fr$ .

A similar discussion holds for the variance. With every  $b$  vertex that we add to the diagram we have to add at least one  $X_0$  or  $D$  vertex which makes it immediately a few orders of magnitude less important. Therefore, it is in good approximation justified to consider only (2.38) for the contribution to the variance.

## 2.4 Empirical estimation of drift and volatility

Based on the derivations of the previous section (by comparing (2.3) with (2.4) and using the results (2.40) and (2.43)) we predict a volatility proportional to  $(s_{\text{eq}} - 1.20)^{3/2}$  and a skewness proportional to  $(s_{\text{eq}} - 1.20)^2$ . We have tested this prediction using various empirical proxies for  $s_{\text{eq}}$  (quantiles and averages), but without any success (essentially pure noise is observed, thus its presentation is omitted). Consequently, we turn to non-parametric methods and test the hypothesis (2.3) by extracting the terms  $f$  and  $g$  directly from the empirical data, using the definition [31]

$$f(s, t) \equiv \lim_{\tau \rightarrow 0} \frac{1}{\tau} \mathbb{E} [s(t + \tau) - s(t)] \quad (2.46a)$$

$$g(s, t) \equiv \sqrt{\lim_{\tau \rightarrow 0} \frac{1}{\tau} \mathbb{E} [(s(t + \tau) - s(t))^2]} \quad (2.46b)$$

with  $\mathbb{E}[\cdot]$  the theoretical expectation operator. Assume that we are given a discrete time series consisting of  $N$  data points  $s_1, s_2, \dots, s_N$ , which is a discrete realisation of a stochastic process  $\{s(t)\}_{t \geq 0}$  (for instance a dataset of historical log exchange rates). The temporal distance between two succeeding data points  $s_i$  and  $s_{i+1}$  is equal to  $\tau$  ( $0 < \tau \ll 1$ ) and assumed independent of  $i$ . Under the additional assumption that the process is stationary,  $f(s, t) = f(s)$ ,  $g(s, t) = g(s)$ , a parameter-free approach to extract  $f$  and  $g$  directly from this time series is obtained by slicing up the value range  $[\min_i s_i, \max_i s_i]$  of the time series into  $K$  bins  $B_\ell$ ,  $\ell = 1, \dots, K$  and

approximating  $f$  and  $g$  in each bin according to [32]

$$f(s^\ell) \approx \frac{1}{\tau} \sum_{i \in B_\ell} (s_{i+1} - s_i) \quad (2.47a)$$

$$g(s^\ell) \approx \sqrt{\frac{1}{\tau} \sum_{i \in B_\ell} (s_{i+1} - s_i)^2}, \quad (2.47b)$$

where  $s^\ell$  denotes the mid point of the  $\ell$ -th bin and the summation is meant over all data points  $s_i$  that lie in the  $\ell$ -th bin. For our application, we download tick by tick data of the EUR/CHF exchange rate, which is then coarse-grained to equally spaced time stamps of 10 seconds ( $\tau = 1/360$  hours) by taking the median. The result is shown in Figure 2.3 for  $K = 100$  bins. Remarkably, we find that  $f$  is essentially constant (and close to 0), in complete contradiction

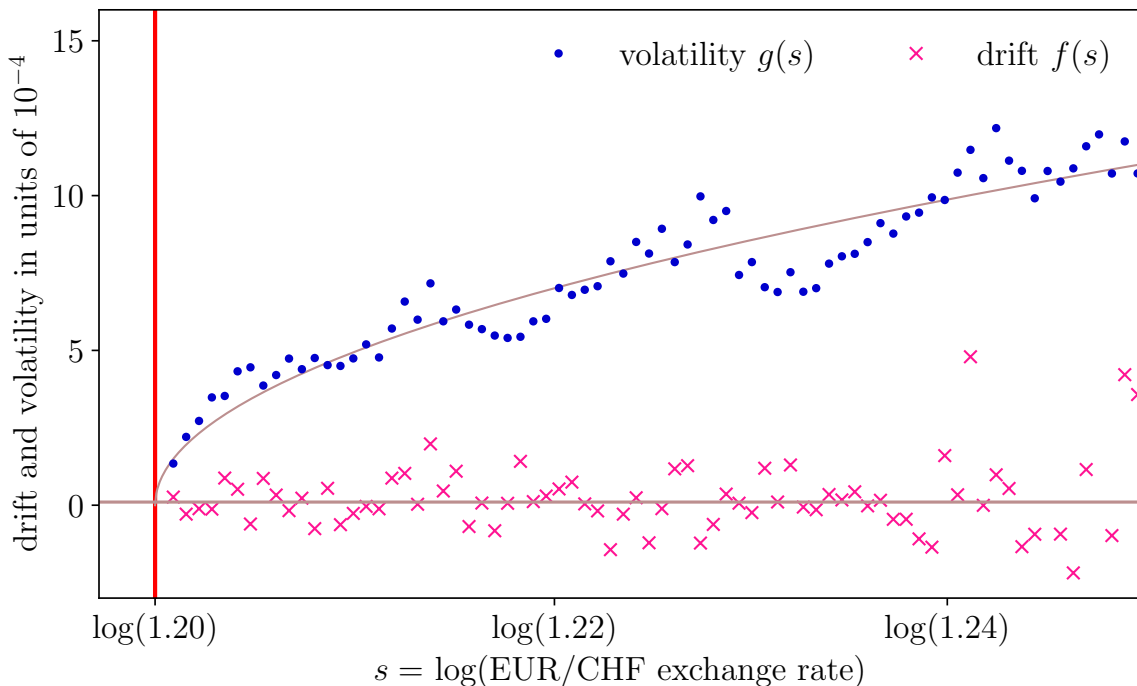


Figure 2.3: We show the parameter free estimate of drift  $f(s)$  and volatility  $g(s)$  obtained from EUR/CHF exchange rate data between September 2011 and January 2015. The brown lines represents the best one-parameter least squares fit to the predicted constant drift and square-root shaped volatility. See also next chapter for an in-depth discussion of this fit.

with the constrained random walk entropic argument: there is no entropic or other potential-derived force acting on the particle. The second interesting observation is that it is  $g$  that exhibits a non trivial  $s$  dependence. It turns out that this non trivial behavior of  $g$  is intrinsic to the target zone regime. We have applied the algorithm of Friedrich et al. [32] to EUR/CHF exchange rate data before September 2011 (Figure 2.4). In the period preceding the committed action of

the Swiss National Bank,  $g$  remains approximately constant over a large range of values, thus recovering the standard GBM model.

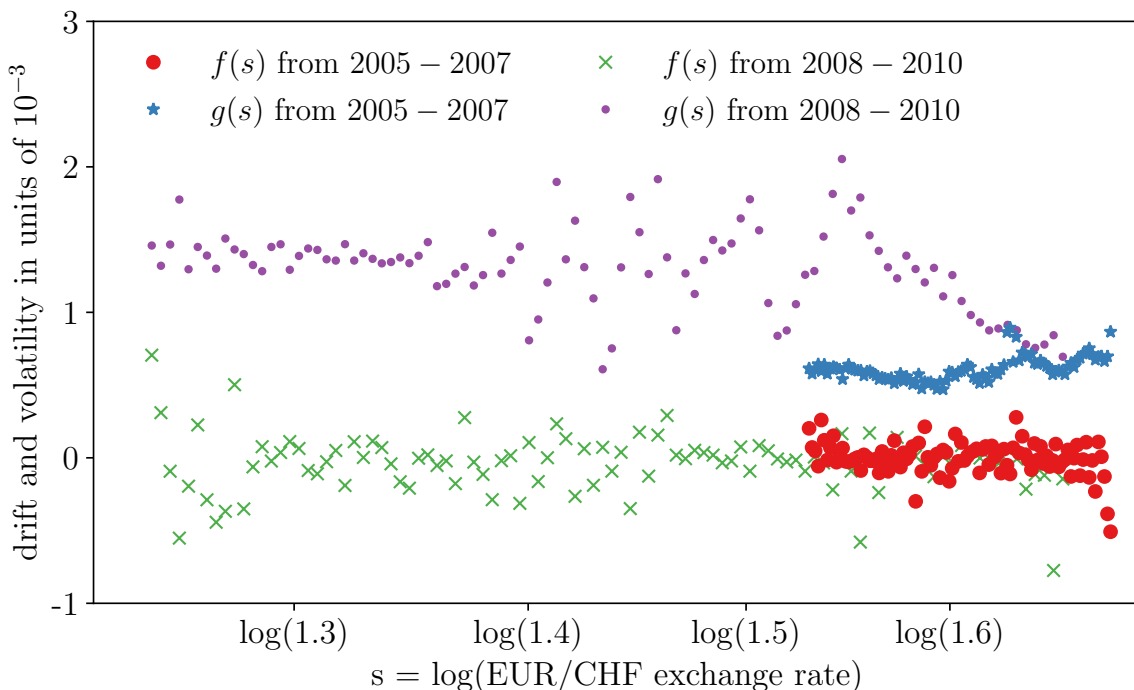


Figure 2.4: Approximation of drift  $f$  and volatility  $g$  using 10 seconds data of the EUR/CHF exchange rate from 2005 to 2007. This figure is obtained by applying the algorithm by Friedrich et al. [32]. In contrast to the target zone regime, we observe that here  $g$  is roughly constant. We have chosen data ranging over periods of three years in order to have approximately the same amount of data points as during the target zone regime.

The above algorithm depends on two parameters, the time step  $\tau$  and the number of bins  $K$ , and their influence on the outcome should be discussed. The dependence on  $K$  is intuitively clear: the larger  $K$ , the larger the bin size and consequently we reduce statistical errors at the price of a less precise resolution. So the choice of  $K$  results from a trade-off. However, as Figure 2.5 confirms, the result is not strongly dependent on the choice of  $K$ .

The approximation (2.47) is only valid if the temporal distance  $\tau$  between two successive data points is small enough. Since  $g$  is a priori unknown, it is difficult to say whether the approximation is justified. To check for the  $\tau$  dependence, we apply approximations (2.47) to the same time series but skip every second data point [33]. Hence, we are testing a time series with time steps  $2\tau$ . This can be repeated iteratively ( $2\tau \rightarrow 4\tau \rightarrow \dots$ ). If the indicated shapes of  $g$  remains stable under several steps of this subsampling, we can deduce that a first order correction term is negligible. Figure 2.5 confirms that this is the case for our application.

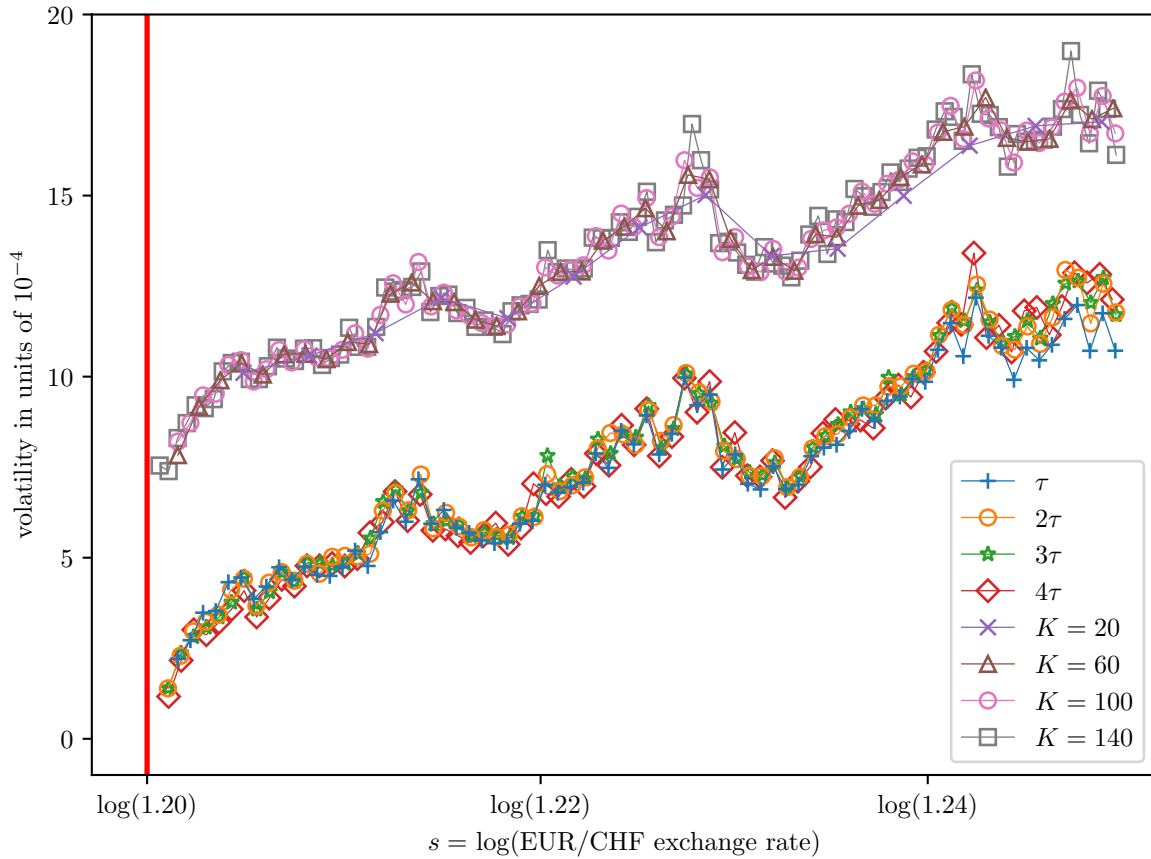


Figure 2.5: Parameter-free estimate of the conditional volatility (2.47b) for different values of  $\tau$  (discrete time step) and  $K$  (number of bin sizes, shown with an offset of  $6 \times 10^{-4}$ ). We see that the indicated shape of  $g(s)$  remains robust over a large range of values for  $K$  and  $\tau$ .

## 2.5 The no-arbitrage condition

The empirical results from Figure 2.3 clearly rejects the simple naive physical model (2.3). Before we go on and look for answers in a more sophisticated physical model, let us consider first what is expected from an economical point of view.

The fact that  $f(s)$  is essentially zero for all  $s$  reveals an important difference between physical and economical Brownian motion. The exchange rate fluctuations are not due to unconscious random actions as would be the myriads of collisions of fluid molecules on a Brownian particle but due to the decisions of investors trying to extract profit from their investments. The aggregate result of this behavior of extremely motivated and driven agents is the quasi-absence of arbitrage, namely the impossibility to extract an excess return. The no-arbitrage condition is one of the organising principles of financial mathematics and is expressed in general by the condition that the process  $s(t)$  obeying (2.1) should be a martingale [5]. In a risk neutral framework (which means that investors do not require additional return for being exposed to risks), this translates

mechanically into the condition of zero drift  $f(s) = 0$ . In the presence of risk aversion, small values of  $f(s)$  are present to remunerate the investors from their expositions to the risks associated with the fluctuating prices. If there was a well-defined, significant drift, knowledge thereof could immediately be translated into sure gains.

To illustrate that this would be the case in our simplified physical model (2.3), we simulated synthetic time series with the generating process (2.3), which has a non-zero  $f(s)$ , with parameters chosen to match the empirical volatility. We used the simple strategy of selling (resp. buying) the euro and buying (resp. selling) the Swiss franc whenever  $s > s_{\text{eq}}$  (resp.  $s < s_{\text{eq}}$ ). Including typical transaction costs between 1 and 2 pips (1 pip = 0.0001 is approximately equal to the bid-ask spread of the real EUR/CHF tick data from Sept. 6, 2011 to January 14, 2015), we find this strategy to deliver extremely high, two-digit annualised Sharpe ratios (as a benchmark, it is typical for mutual funds, hedge-funds and the market portfolio itself to deliver performances with Sharpe ratios less than 1, and often much less than 1). This clearly illustrates that the process (2.3) would lead to exchange rates that can be forecasted, which would ‘leave enormous amount of money on the table’. It is thus completely unrealistic from a financial view point.

Instead, Figure 2.3 suggests that the drift is constant,  $f \equiv c \approx 0$ , and the volatility has a non-analytic shape  $g(s) \sim (s - \underline{s})^\gamma$  for some  $\gamma \in (0, 1)$ . In the next section, we will show that  $\gamma = 1/2$  is the natural choice from a physical perspective. In the subsequent chapter, we will furthermore see that  $\gamma = 1/2$  is also the prediction from the classical economics literature.

## 2.6 A physicist’s second approach

We have shown in the previous section that the initial naive physical model failed to account to for the action anticipating traders. In the next chapter, we will see how this ingredient is fundamentally incorporated within the economics literature. In broader terms, we can classify this finding as ‘the presence of the wall must be felt even away from the wall’ at all times and for all random walk realisations. We use this insight to develop a second, more elaborate physical model. We note that the physical Brownian particle can receive the information ‘there is a wall’ not only through direct contact but also through the fluid.

Consider a physical Brownian particle in a fluid. The presence of a wall leads to a modification of the hydrodynamic flow of the molecules trapped between the wall and the Brownian particle. The closer the Brownian particle to the wall, the thinner the lubrication layer between them

and the more hindered is the diffusion of the Brownian particle. In physics, it is more common to work with the diffusion coefficient  $D(s)$  which is related to our volatility via  $g = \sqrt{2D}$ . In the bulk of a fluid (where the wall is not felt), the diffusion coefficient  $D$  is a constant  $D_0$ . The Einstein-Stokes equation predicts for a spherical particle with radius  $R$

$$D_0 = \frac{k_B T}{6\pi\nu R} \quad (2.48)$$

with  $k_B$  the Boltzmann constant,  $T$  the temperature and  $\nu$  the viscosity of the fluid. In presence of a wall at  $s = \underline{s}$ , the diffusion coefficient must be modified by  $D(s) = D_0/\lambda$  where  $\lambda$  is given by [34]

$$\lambda = \frac{4}{3} \sinh(\alpha) \sum_{n=1}^{\infty} \left[ \frac{n(n+1)}{(2n-1)(2n+3)} \frac{2 \sinh((n-1)\alpha) + (2n+1) \sinh(\alpha)}{4 \sinh^2((n+1/2)\alpha) - (2n-1)^2 \sinh^2(2\alpha)} \right] \quad (2.49)$$

and  $\alpha = \text{arcosh}(s/R)$ . The resulting shape is depicted in Figure 2.6 and shows that the presence of a barrier translates into the decrease due to hydrodynamic forces of the diffusion coefficient of the Brownian particle upon its approach to the wall. An approximation of this result had

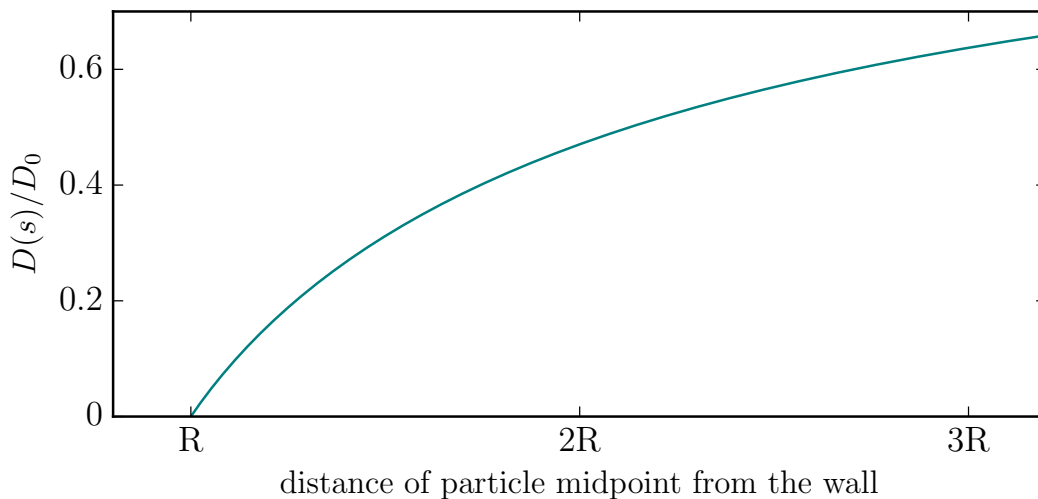


Figure 2.6: Physical diffusion coefficient as a function of particle distance from the wall. To first order and close to the wall,  $D(s)$  is a linear function of  $s - \underline{s}$ .

already been found decades before by Lorentz [35] who predicted

$$\lambda \sim 1 + \frac{9}{8} \frac{R}{s - \underline{s}} \quad (2.50)$$

from which we find, to first order,  $D(s) \sim (s - \underline{s})$ . It follows immediately that, close to the wall,

$$g(s) = \beta\sqrt{s - \underline{s}} \quad (2.51)$$

for some constant  $\beta > 0$ . In the next chapter, the same prediction is derived from a completely different perspective. This shows that there is an analogy between physical and economical Brownian motion after all. Had we relied from the beginning on the correct physical picture of hindered diffusion, we could have predicted the behavior of the exchange rate close to the barrier solely on the basis of physical theories.

The analogy that we have found here extends a previous result [6] in which it has been shown that the GBM model of financial price fluctuations is deeply anchored in the physics-finance analogy of a colloidal Brownian particle embedded in a fluid of molecules as shown in Figure 2.2 (omitting the previously shown incorrect potentials), where the surrounding molecules reflect the structure of the underlying order book. Here, we have shown that this analogy holds even if the order book fluid is restricted from one side by an upper or lower bound.

## 2.7 Equilibrium in physics and in economics

In absence of any external force, what is the stochastic process that describes a physical Brownian particle? Naively, one is led to propose  $ds/dt = g(s) \eta(t)$ . However, this implies not only that we are working in Itô's interpretation of stochastic calculus, but can furthermore be shown to be inconsistent with convergence towards thermal equilibrium. For a system at equilibrium, the probability density  $p(s, t)$  must have a steady state solution with the canonical form  $p(s) \sim \exp(-\mathcal{H}/k_B T)$  with  $\mathcal{H}$  the Hamiltonian of the system. If we insist on working in Itô's interpretation as is customary in finance to ensure causality of financial strategies, one must include an additional drift term  $g(s) \frac{dg(s)}{ds}$  to the stochastic differential equation in order to be consistent with the physical steady state distribution (see section 2.2.3 of [36] and [37] for derivations). From the volatility (2.51), we then derive the following stochastic equation for a Brownian particle close to a wall and in absence of external forces:

$$\frac{ds}{dt} = g(s) \frac{dg(s)}{ds} + g(s) \eta(t) = \frac{\beta^2}{2} + \beta\sqrt{s - \underline{s}} \eta(t). \quad (2.52)$$



Remarkably, the square-root shaped volatility is exactly the function that induces a constant positive drift in agreement the empirical observation in figure 2.3, and in agreement with the prediction from the economics literature, that will be discussed in the subsequent chapter. From a purely physical perspective, this result (2.52) has another interesting implication. It reveals the special role played by the linearly increasing diffusion coefficient. As the following simple heuristic derivation shows, a locally linear diffusion coefficient is the only physically sensible choice, which can be predicted in absence of any detailed knowledge about hydrodynamic interactions. We start from the fairly general assumption that, close to the wall, without loss of generality located at  $\underline{s} = 0$ , the volatility is of the form  $g(s) = \beta s^\gamma$  for some  $\gamma > 0$  (it is easy to see that  $\lim_{s \downarrow 0} D(s) = 0$  is a necessary condition). Plugging the ansatz into (2.52) gives

$$\frac{ds}{dt} = \beta^2 \gamma s^{2\gamma-1} + \beta s^\gamma \eta(t). \quad (2.53)$$

In the limit  $s \downarrow 0$ , we can distinguish three cases:

$$\text{the drift } g(s) \frac{dg(s)}{ds} \begin{cases} \text{diverges} & \text{if } \gamma < 1/2, \\ \text{is constant} & \text{if } \gamma = 1/2, \\ \text{vanishes} & \text{if } \gamma > 1/2. \end{cases}$$

If  $\gamma < 1/2$ , the particle will be repelled with infinite force and can never touch the wall. Furthermore, placing initially the particle at the wall is ill-defined. If  $\gamma > 1/2$ , the particle, once it has reached the wall, will stay there forever (more precisely, it can be shown that a particle starting from  $s > 0$  can never exactly reach the wall, but approach it arbitrarily close [38]). Also, a particle placed at the the wall will simply stay there forever. We deduce that  $\gamma = 1/2$ , and hence  $D(s) \sim s$  is the only physically reasonable choice. In this case, a particle starting from  $s \geq 0$  has non-zero probability to reach the boundary in finite time, upon which it will be repelled. These arguments can be formalised by solving analytically the Fokker-Planck equation corresponding to (2.53) in terms of an eigenfunction expansion [31, 39].

## 2.8 Conclusions

In a first attempt, we have approached the problem of currency target zones from a statistical physics perspective, describing its dynamics in terms of a free random walk and a hard repelling barrier. The predicted entropic potential implies a deterministic drift that turns out completely

unrealistic from a financial point of view, violating the no-arbitrage principle. In contrast to the trajectory of a physical particle, a financial price time series is not governed by purely mechanical laws, but actively embodies the information of the presence of a constraint, which is fed back into the price itself. By measuring drift and volatility of the stochastic price process non-parametrically, we could show that indeed the drift term is vanishing, whereas the volatility displays a non-trivial square-root decay as a function of the distance to the barrier. By describing the exchange rate as a colloidal Brownian particle embedded in an ‘order book fluid’, we could derive that square root shape from first order principles. This provides novel empirical support for the recently introduced model of a ‘financial Brownian particle in a layered order book fluid’ generalizing the standard random walk model of financial price fluctuations. From a didactical perspective, we have given a complete example of how economic models can be motivated by ideas and mathematical tools from physics and vice versa. We have also pointed out a fundamental difference between physical and economic hindered diffusion. In physics, we have an additional constraint in terms of a thermal equilibrium based on detailed balance. In finance, the existence of such a global equilibrium is a priori not clear and must be investigated further. It would not be unsurprising, if further motivations or analogies can be found from the study of physical out-of-equilibrium systems.

## Chapter 3

# Quantitative modelling of the EUR/CHF exchange rate during the target zone regime of September 2011 to January 2015 <sup>†</sup>

In the previous chapter, we have discussed economic currency target zones from a physical perspective. But not surprisingly, currency target zones have been under heavy scrutiny in the economics literature. A landmark paper is by Krugman [40], known for its simplicity and elegance. Krugman assumes that (i) the target zone is perfectly credible (market participants hold no doubt that the exchange rate will not be let to move outside the band), (ii) the central bank intervenes only (infinitesimally) close to the boundaries (so called marginal intervention) and (iii) the target zone is exactly defended (no exchange rates occur outside the target zone band). <sup>1</sup> While empirical evidence has been sparse so far (see section 3.1), this model has remained the benchmark model for a large part of the target zone literature. In order to account for the various reported deviations from Krugman's assumptions or predictions, several extensions have been proposed.

For instance, it has been shown that the assumption of perfect credibility do not hold for the EMS era [41]. A variety of different approaches have been suggested to adjust for this fact [42–45],

---

<sup>†</sup>This chapter is based on reference [10].

<sup>1</sup> This last assumption is actually just a consequence of the first one. However, one could think of a target zone that is perfectly credible in the sense that it is always defended by the central bank, but is allowed to have tiny fluctuations around its limits. See also the discussion of soft target zones below.

mostly by including a probability factor that incorporates the possibility that the target zone is readjusted or given up entirely.

While, in theory, the central bank is expected to wait idle until the exchange rate approaches very close to its limits when it acts, this is not what is often observed in practice. Rather, the central bank will not refrain from interacting preemptively even when the exchange rate is well within the band, thus violating the assumption of marginal intervention [46–50].

Krugman's model has also been extended to multilateral target zones [51, 52]. It is found that several key predictions from Krugman's model (honeymoon effect and U-shaped density distribution, see section 3.1) may no longer hold when taking into consideration the mutual interdependencies between different currency pairs.

Instead of directly targeting the spot exchange rate, target zones for the forward exchange rate to be delivered at a maturity date have been suggested [53]. To defend the forward-rate target zone, the central bank needs to undertake forward sales or purchases of foreign exchange when the target forward rate is to escape a specified band. While the model describes similar (yet diminished) stabilizing effects on the spot exchange rate compared with the Krugman spot-rate target zone, the incidence on the central bank's foreign reserves in the futures model is not only delayed, but also smaller. This may reduce the probability of the target zone to collapse.

Another extension of Krugman's model allows for so-called soft target zones [54, 55]. Assuming that it is not the exchange rate but rather its moving average that is confined within the target zone boundaries, such a model accounts for the small fluctuations that are sometimes observed in target zone arrangements. An advantage, from the central bank point of view, of allowing small fluctuations is to hinder speculative attacks. A recent example of this type is the target zone for the US dollar/Hong-Kong dollar (USD/HKD) exchange rate imposed by the Hong-Kong Monetary Authority (HKMA) in 2005, that will be discussed in more depth in the next chapter. See also refs. [56, 57] for USD/HKD specific investigations. For an exhaustive overview of other approaches that have been suggested, see ref. [58].

Our first contribution to the literature is to propose new metrics that can be readily estimated empirically and which benefit from unambiguous predictions of the Krugman target zone model, including a direct test for the smooth pasting condition, or equivalently, the no-arbitrage condition. Our second contribution is to use these metrics to what we claim is arguably the cleanest empirical case of a credible target zone, namely the case of the euro/Swiss franc exchange rate in the extraordinary period from September 6, 2011, to January 15, 2015, when the Swiss National

Bank enforced a minimum exchange rate of 1.20 Swiss francs per euro. The comparison between the normative predictions of Krugman's model and their empirical counterparts allows us to test Krugman's model in a clean context in which Krugman's assumptions apply.

The remainder of this chapter is organized as follows. In section 3.1, we present Krugman's target zone model and in section 3.2 we review the related empirical literature, which happens to have mainly rejected the model. We derive expressions for the conditional volatility and for the probability density of the exchange rate in section 3.3. This enables us to test Krugman's predictions directly from observable data. This is what we do in section 3.4 by calibrating these expressions to the euro/Swiss franc exchange rate data from September 2011 to January 2015. Section 3.5 concludes.

### 3.1 Krugman's target zone model

The work of Krugman [40] has become a reference of a large part of the target zone literature. Following refs. [40, 59], we summarise now his model, starting from the following intuitive equation:

$$s = m + v + \gamma \frac{\mathbb{E}[ds]}{dt}. \quad (3.1)$$

Here,  $s$  is the (natural logarithm of) the foreign exchange spot price, also called the exchange rate (the logarithm is implicit, in what follows). The exchange rate is the domestic price of the foreign currency, that is, the number of domestic currency units needed to buy one unit of foreign currency. In a target zone,  $s$  is held inside the target zone band  $\underline{s} < s < \bar{s}$ . We call  $\underline{s}$  the strong side of the target zone band, because the smaller  $s$ , the less domestic money one needs to buy one unit of the foreign currency. Analogously,  $\bar{s}$  is called the weak side of the band. For the case of the euro/Swiss franc target zone examined in this chapter, only the strong limit  $\underline{s} = \log(1.20$  Swiss francs per euro) is imposed. In the chapter, we will mostly ignore the upper limit  $\bar{s}$ , or formally, set  $\bar{s} \rightarrow \infty$ . Since we are only interested in the behavior of  $s$  close to  $\underline{s}$ , one verifies easily that this does not change anything about the qualitative and quantitative predictions of the model. The quantity  $m$  denotes the (natural logarithm of) the domestic money supply, which is supposed to be controlled by the central bank. When the currency is strong (the exchange rate is low), the central bank can increase the money supply in order to weaken the domestic currency. Krugman assumes that the monetary policy is passive, i.e.  $m$  is shifted only at (or infinitesimally close to) the exchange rate limit  $\underline{s}$ . As long as  $s$  is well above  $\underline{s}$ ,  $m$  is held constant.

This is usually referred to as marginal interventions. By  $v$ , we denote (the natural logarithm of) exogenous velocity shocks, i.e. economic and geopolitical events affecting the exchange rate that cannot be controlled by the (domestic) central bank. The velocity term is assumed to follow a simple Brownian motion

$$dv = \sigma dW_t \quad (3.2)$$

with  $\sigma > 0$  constant. We call  $f \equiv m + v$  the fundamental value of the exchange rate, i.e. the value in absence of any regulatory restrictions. The final, crucial ingredient to the model is the expected change of the exchange rate  $\mathbb{E}[ds]/dt$ . This is the component that takes into account the psychology of market participants, and exactly what has been missing in the mechanical considerations of the previous chapter. In a credible target zone, when the exchange rate approaches  $\underline{s}$ , market participants anticipate the intervention by the central bank and act accordingly. The constant  $\gamma > 0$  can be interpreted as the semi-elasticity of the exchange rate with respect to the instantaneous expected rate of currency depreciation. For now, we just treat it as a constant and derive a deeper understanding in chapter 5.

In order to proceed with model (3.1), we must somehow rewrite the expectation term  $\mathbb{E}[ds]/dt$ . To this end, we apply Itô's lemma to  $s(f)$ , and take the expectation value,

$$\mathbb{E}[ds] = \mathbb{E}\left[\frac{\sigma^2}{2} \frac{\partial^2 s}{\partial f^2} dt\right] + \mathbb{E}\left[\sigma \frac{\partial s}{\partial f} dW_t\right] = \frac{\sigma^2}{2} \frac{\partial^2 s}{\partial f^2} dt. \quad (3.3)$$

Plugging (3.3) into (3.1) yields

$$s = f + \frac{\gamma \sigma^2}{2} \frac{\partial^2 s}{\partial f^2}. \quad (3.4)$$

This simple second order linear ordinary differential equation has the general solution

$$s(f) = f + Ae^{-\rho f} + Be^{\rho f} \quad (3.5)$$

with

$$\rho \equiv \sqrt{\frac{2}{\gamma \sigma^2}} \quad (3.6)$$

and constants  $A, B$ . In the limit far above the lower limit  $\underline{s}$ , where the presence of the barrier is not felt, we must recover  $s = f$ . We conclude that  $B = 0$ . Denote by  $\underline{f}$  the value at which  $s(\underline{f}) = \underline{s}$ . This must be understood in the limit  $\lim_{f \downarrow \underline{f}} s(f)$  because as soon as  $s$  touches  $\underline{s}$ ,  $m$  will be changed, making it difficult to define a unique value  $\underline{f}$ . The constant  $A$  is determined by

the smooth-pasting condition

$$\left. \frac{\partial s}{\partial f} \right|_{f=\underline{f}} = 0, \quad (3.7)$$

which is based on a no-arbitrage argument [58, 60]. To see this, consider a point  $f^* \gtrsim \underline{f}$  close to  $\underline{f}$  at which the authorities intervene by increasing the money supply in such a way that  $f$  is pushed back into the band by a discrete amount  $\Delta f$ . Hence, whenever the exchange rate gets to  $s(f^*)$ , the authorities intervene by a discrete amount, resulting in the exchange rate becoming instantly equal to  $s(f^* + \Delta f)$ . As soon as  $s$  touches  $s(f^*)$ , then it would appear that the traders might enjoy a risk-free opportunity, since they could buy the foreign currency just before the intervention and make a profit equal to the difference between  $s(f^*)$  and  $s(f^* + \Delta f)$ . This arbitrage opportunity is only eliminated if

$$s(f^*) = s(f^* + \Delta f). \quad (3.8)$$

By assumption, the central bank does not intervene at a point  $f^* > \underline{f}$  but only infinitesimally close to  $\underline{f}$ . Hence, dividing (3.8) by  $\Delta f$  and taking the infinitesimal limit we find

$$\lim_{f^* \downarrow \underline{f}} \lim_{\Delta f \downarrow 0} \frac{s(f^* + \Delta f) - s(f^*)}{\Delta f} = 0. \quad (3.9)$$

Since  $s$  is a continuous function of  $f$ , this later expression is exactly the imposed smooth-pasting condition (3.7). Hence,  $A = \frac{1}{\rho} e^{\rho \underline{f}}$  and

$$s(f) = f + \frac{1}{\rho} e^{\rho(\underline{f}-f)}, \quad (3.10)$$

which completes the derivation of the model. The result is depicted in Figure 3.1.<sup>2</sup> It is clearly visible that the closer  $s$  to the lower limit  $\underline{s}$ , the less it reacts to external shocks  $v$ . Since this happens without active intervention of the central bank (which is assumed to act only at the limit at which  $s$  touches  $\underline{s}$ ), the sole announcement of a target zone renders the exchange rate less volatile. This is known as the honeymoon effect.<sup>3</sup>

Extending the above derivation to two-sided target zones, and a volatility  $v$  with a drift term,

<sup>2</sup>For the two-sided target zone, (3.10) is modified to  $s(f) = f - A \sinh(\rho f)$  for some positive constant  $A$ , determined from smooth pasting. This leads to an S-shaped relation between  $s$  and  $f$ , see for instance Figure 5 in [40].

<sup>3</sup>This effect may be reinforced by setting up a separate forward-rate target zone [53], or weakened by having a multilateral target zone [51, 52].

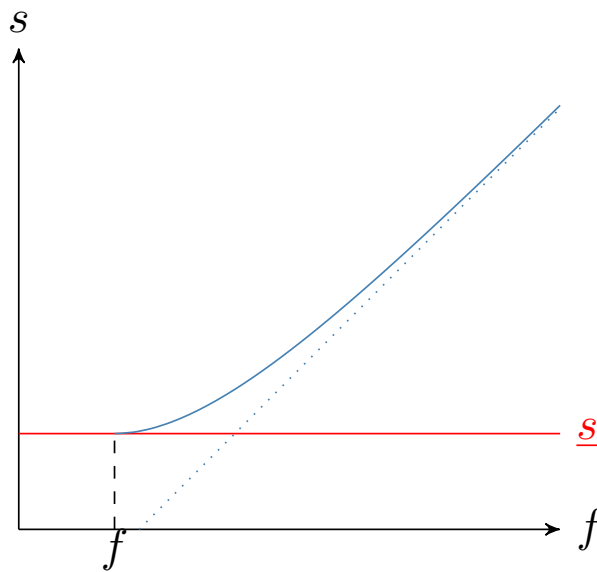


Figure 3.1: Schematic plot of Krugman's exchange rate model (3.10) showing the exchange rate  $s$  as a function of the fundamental  $f$ , bounded from below by  $\underline{s}$ . The smooth-pasting condition (3.7) accounting for the no-arbitrage condition is clearly visible. The dashed line shows the trivial relation  $s(f) = f$  that prevails under a freely floating exchange rate polity.

$dv = \mu dt + \sigma dW_t$  is similar, resulting in the solution (c.f. e.g. ref. [58]),

$$s = f + A \exp(\xi_1 f) + B \exp(\xi_2 f) \quad (3.11)$$

where  $\xi_1, \xi_2$  are the roots of the quadratic equation  $\xi^2 \gamma \sigma^2 + 2\xi \gamma \mu = 2$ . Both  $A$  and  $B$  are now fixed through the two smooth pasting condition at  $\underline{f}$  and  $\bar{f}$ , respectively. See Figure 3.2 for examples.

## 3.2 Previous empirical tests of Krugman's model

Following refs. [50, 58, 59], we summarise here what the empirical research reports about Krugman's target zone model. Results for the euro/Swiss franc target zone investigated by [61] are omitted here and will be presented in section 3.4.

Krugman's model is based on two crucial assumptions. First, the exchange rate is perfectly credible, i.e. all market participants are convinced that the target zone will be defended. Second, the central bank intervenes only marginally, i.e. "infinitesimally" close to the barrier. Many studies have challenged the model by questioning the validity of these restrictive assumptions.

One way to test credibility is by examining whether forward exchange rates for different maturities fall outside the exchange rate band [62]. If this is the case for some maturities, the



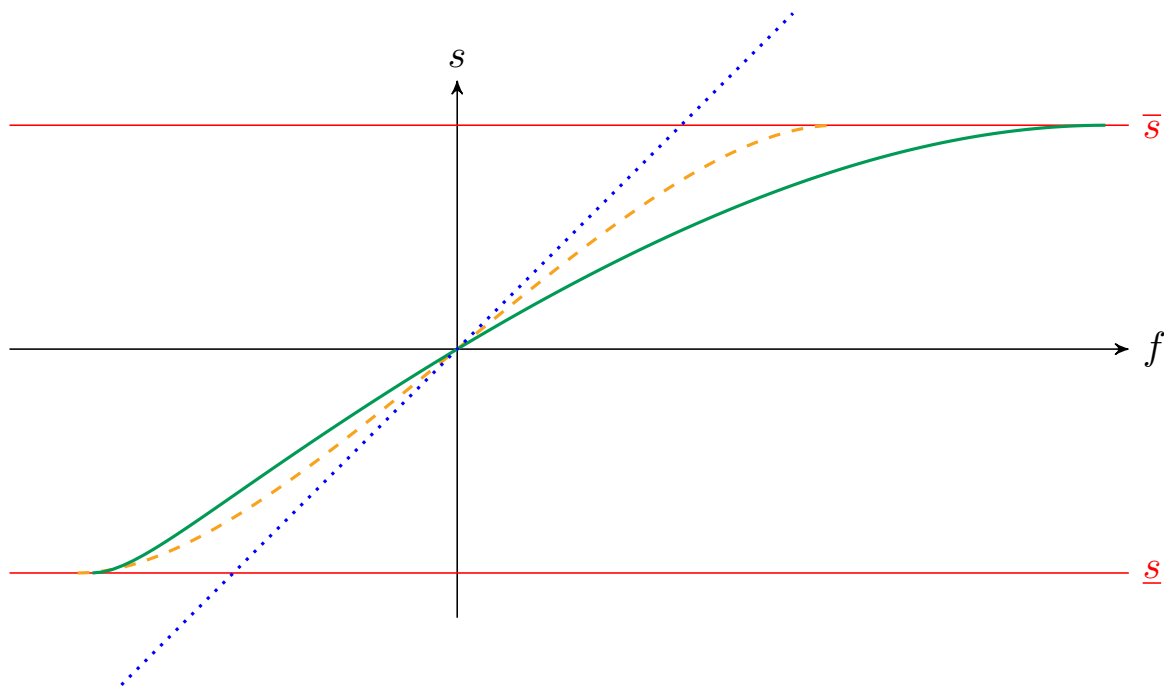


Figure 3.2: Two solutions of the two-sided Krugman target zone model (3.11). The dashed line corresponds to  $\mu = 0, \sigma = 1$  and  $\gamma = 1$ . The absence of a drift in the fundamental makes this solution symmetric. The second solution (continuous line) has parameters  $\mu = 3, \sigma = 1$  and  $\gamma = 1$  and is therefore not symmetric. Without loss of generality, we have shifted the solutions such that  $s(f = 0) = 0$ . The dotted line represents the relationship  $s = f$  that holds in absence of a target zone due to market efficiency.

exchange rate band cannot be perfectly credible. Using various statistical tests, the assumption of perfect credibility is generally rejected [62, 63].

The assumption of marginal interventions is also strongly rejected by the empirical literature. Most tellingly, marginal interventions appear to represent the exception rather than the rule followed by central banks [64–66].

The most important implication of the model is that the exchange rate should display a distinct non-linear relationship between the fundamental  $f$  and the exchange rate  $s$  as predicted by (3.10). Plots of the exchange rate against the fundamentals (variously defined), as well as more sophisticated tests, rarely reveal anything resembling the predicted shape. Using weekly data for the French franc and the German mark against the US dollar for the sample period 1899–1909 and daily data for the French franc and the Danish krone against the German mark for the sample period March 14, 1979 to March 1, 1989, no out-of-sample evidence for Krugman's model has been found [67]. Using daily data for all the EMS countries and Euro-interest rates during the sample period March 30, 1979 to May 16, 1990 also leads to rejection of the model [63]. Testing the target zone model on data for the German mark-Italian lira exchange rate from January 14,

1987 to September 20, 1989 provides further evidence against the model [68]. Similarly for a number of other EMS currencies [69–71]. Some supportive evidence for the basic target zone model is found for daily data for the Swedish krona and Euro-interest rates from January 4, 1982 to November 15, 1990 [66].

Other implications of the target zone model, such as the  $\cup$ -shaped probability density <sup>4</sup> are strongly rejected by the empirical evidence which, by contrast, suggests a hump-shaped distribution function for exchange rates operating under target zone arrangements, with most of the probability mass lying in the interior of the band and only a small portion near the edges of the band [43, 63, 66].

### 3.3 Two new observables to test Krugman's model

We have seen in the previous section that empirical evidence for Krugman's target zone model has been lacking. The model has often been rejected just on the basis of its assumptions. But even for the exemplary case such of the EUR/CHF target zone where the target zone could be arguably assumed 'perfectly' credible and only marginally defended, only a noisy relation between observable exchange rate and estimated fundamental has been reported [61] (see also section 3.4). This motivates us to propose a different approach. Rather than relying on an ambiguous estimate of the fundamental  $f$ , we invert locally the relation between the exchange rate  $s$  and  $f$  and express the conditional volatility  $g(s)$  and probability density  $p(s)$  directly as a function of the observable exchange rate  $s$ .

Evidently, the function  $s : [\underline{f}, \infty] \rightarrow [\underline{s}, \infty]$  given by (3.10) has a well-defined global inverse, which, however, has no analytical closed form expression. Since we are only interested in the relation between  $s$  and  $f$  close to the lower barrier  $\underline{s}$ , we consider the Taylor approximation

$$\begin{aligned} s(f) &= \underbrace{s(\underline{f})}_{=\underline{s}} + \underbrace{s'(\underline{f})(f - \underline{f})}_{=0} + \frac{1}{2}s''(\underline{f})(f - \underline{f})^2 + \mathcal{O}\left((f - \underline{f})^3\right) \\ &= \underline{s} + \frac{1}{2}\rho(f - \underline{f})^2 + \mathcal{O}\left((f - \underline{f})^3\right) \\ &\approx \underline{s} + \frac{1}{2}\rho(f - \underline{f})^2. \end{aligned} \tag{3.12}$$

It is not clear, a priori, that this Taylor approximation is allowed. Indeed, (3.12) requires case-specific justification, in particular because  $f$ , defined via (3.2), can fluctuate in principle between

---

<sup>4</sup> $\cup$ -shaped for the two-sided target zone, as is clear from symmetry.

$-\infty$  and  $+\infty$ . We will show in section 3.4 that the approximation holds very well for the EUR/CHF case. Relation (3.12) can now easily be inverted:

$$f = \sqrt{\frac{2}{\rho}} \sqrt{s - \underline{s}} + \underline{f}. \quad (3.13)$$

The derivation works equally well even if we do not insist on the smooth pasting condition  $s(\underline{f}) = 0$ . We will make use of this remark in section 3.4, when we test the validity of this assumption explicitly for the euro/Swiss franc target zone.

### 3.3.1 Derivation of the conditional volatility

Our first observable is the conditional volatility  $g$ . Applying Itô's lemma to (3.10) gives the following expression for the volatility:

$$g(f) = \sigma \frac{ds}{df} = \sigma - \sigma e^{\rho(\underline{f}-f)}. \quad (3.14)$$

Plugging expansion (3.13) into (3.14) gives to leading order:

$$\begin{aligned} g(s) &= \sigma - \sigma e^{\rho \underline{f}} \exp(-\rho f) \\ &= \sigma - \sigma e^{\rho \underline{f}} \exp\left(-\rho \left[\sqrt{\frac{2}{\rho}} \sqrt{s - \underline{s}} + \underline{f}\right] + \mathcal{O}\left((f - \underline{f})^3\right)\right) \\ &\approx \sigma - \sigma e^{\rho \underline{f}} \exp\left(-\rho \left[\sqrt{\frac{2}{\rho}} \sqrt{s - \underline{s}} + \underline{f}\right]\right) \\ &= \sigma - \sigma \exp\left(-\sqrt{2\rho} \sqrt{s - \underline{s}}\right) \\ &= \sigma - \sigma \left[1 - \sqrt{2\rho} \sqrt{s - \underline{s}} + \rho(s - \underline{s}) + \mathcal{O}\left((s - \underline{s})^{3/2}\right)\right] \\ &\approx \sigma - \sigma \left[1 - \sqrt{2\rho} \sqrt{s - \underline{s}} + \rho(s - \underline{s})\right] \\ &= \sqrt{2\rho} \sigma \sqrt{s - \underline{s}} - \rho \sigma (s - \underline{s}) \end{aligned} \quad (3.15)$$

$$\begin{aligned} &\approx \sqrt{2\rho} \sigma \sqrt{s - \underline{s}} \\ &= \beta \sqrt{s - \underline{s}} \end{aligned} \quad (3.16)$$

where we define

$$\beta \equiv \sqrt{2\rho} \sigma = 2^{1/4} \gamma^{-1/4} \sigma^{1/2} > 0. \quad (3.17)$$

With (3.16), we have an expression that can be tested directly from exchange rate data. Interestingly, this is the same as expression (2.51), derived in the previous chapter from physical

principles.

Expression (3.16) provides a key prediction to test Krugman's model. Indeed, recall that the key insight of Krugman's solution (3.10) is the impact on the exchange rate  $s(f)$  far from  $\underline{s}$  (and thus  $\underline{f}$ ) due to the credible declared intention for action of the central bank, which translates technically into a vanishing of the first-order derivative  $\frac{\partial s}{\partial \underline{f}} = 0$ , i.e. the smooth-pasting condition (3.7). This implies that the function  $s(f)$  is a quadratic function of  $f$  close to  $\underline{f}$  with zero slope. This is the mathematical embodiment of the credibility of the declared intention for action of the central bank on the exchange rate. We should stress that, given that the smooth-pasting condition (3.7) holds and, in absence of other conditions, the function  $s(f)$  has to be a quadratic function of  $f$  close to  $\underline{f}$ , and not for instance a cubic or quartic function of  $f$ . It is this condition of being a quadratic function that leads to the exponent 1/2 in the dependence of the volatility  $g(s)$  given by expression (3.16). If a cubic (respectively quartic) dependence held, the exponent would have been 1/3 (respectively 1/4). Thus, the exponent 1/2 in expression (3.16) shows the universality of the mechanism for the vanishing of the volatility close to the barrier, which is associated with the single constraint of smooth pasting. This justifies that testing for the square-root dependence (3.16) amounts to testing an important component of Krugman's model.

### 3.3.2 Derivation of the probability density of exchange rates

Here, we derive the  $\cup$ -shaped probability density function of the exchange rate  $s$  close to the lower barrier. Note that the unconditional probability density function  $\tilde{p}$  of the fundamental is (asymptotically) uniform close to the barrier,  $\tilde{p}(f) = \tilde{p} = \text{const}$  [66, 72]. Applying the usual rules for a change of variables in a probability distribution, we find for the probability density  $p$  of the exchange rate  $s$

$$p(s) = \tilde{p}(f) \left| \frac{df(s)}{ds} \right| \simeq \tilde{p} \left| \frac{df(s)}{ds} \right|, \quad (3.18)$$

where the sign  $\simeq$  of the second term in the r.h.s uses the asymptotic value  $\tilde{p}(f) = \tilde{p}$  close to the barrier. For  $s$  close to  $\underline{s}$ , we can use once again the approximative identity (3.13) (see section 3.4.6 for a justification) and obtain

$$p(s) = \frac{C}{\sqrt{s - \underline{s}}} \quad (3.19)$$

with constant

$$C = \tilde{p} / \sqrt{2\rho}. \quad (3.20)$$

The interpretation of this result is also clear from (3.16). The closer the exchange rate to  $\underline{s}$ , the smaller the volatility ('the average step size') and hence the more time is spent in regions close to  $\underline{s}$ . In the next section, we compare predictions (3.16) and (3.19) to empirical data.

## 3.4 Application to the euro/Swiss franc target zone

### 3.4.1 History of the euro/Swiss franc target zone

The Swiss franc has a long reputation as being a safe haven currency, meaning a relatively stable currency in a politically and economically well-developed, secure state. When the euro was introduced in 1999, its value was set to 1.58 Swiss francs. In the subsequent years, the exchange rate experienced mild fluctuation with a low of 1.44 on September 2001 and a peak of 1.67 CHF on October 2007, as determined by the free market and guided by a stable European economy and the belief that the creation of the euro zone has led to prosperity for the entire eurozone. This was changed when the European debt crisis made investors lose faith in the European currency union and seek investments in more stable currencies, among them the Swiss franc, thereby increasing demand and strengthening the CHF. The Swiss economy suffered from these developments between 2008 and 2011 since it is largely depending on its exports, of which most of them go to countries in the eurozone. On September 2011, when the EUR/CHF parity was approaching an incredibly low exchange rate of one Swiss franc per euro, passive monetary policies (e.g. interest rates) were no longer effective and the Swiss National Bank (SNB) decided to intervene actively in the market. On the September 6, 2011, the SNB stated in a press release:<sup>5</sup> 'The current massive overvaluation of the Swiss franc poses an acute threat to the Swiss economy and carries the risk of deflationary development. The Swiss National Bank is therefore aiming for a substantial and sustained weakening of the Swiss franc. With immediate effect, it will no longer tolerate a EUR/CHF exchange rate below the minimum rate of CHF 1.20. The SNB will enforce this minimum rate with the utmost determination and is prepared to buy foreign currency in unlimited quantities.' The SNB held on to this policy without any exception for almost three and a half years. Completely unexpected for many market participants and during times of consistent pressure on the Swiss franc, the SNB stated on January 15, 2015,<sup>6</sup> 'The Swiss National Bank is discontinuing the minimum exchange rate of CHF 1.20 per euro. [...]' The minimum exchange rate was introduced during a period of exceptional overvaluation of the

---

<sup>5</sup> [http://www.snb.ch/en/mmr/reference/pre\\_20110906/source/pre\\_20110906.en.pdf](http://www.snb.ch/en/mmr/reference/pre_20110906/source/pre_20110906.en.pdf)

<sup>6</sup> [http://www.snb.ch/en/mmr/reference/pre\\_20150115/source/pre\\_20150115.en.pdf](http://www.snb.ch/en/mmr/reference/pre_20150115/source/pre_20150115.en.pdf)

Swiss franc and an extremely high level of uncertainty on the financial markets. This exceptional and temporary measure protected the Swiss economy from serious harm. [...] The economy was able to take advantage of this phase to adjust to the new situation.' The consequences were immediate and severe. In just one day, the exchange rate dropped from 1.2 to 1.05 CHF per euro and the Swiss Market Index (SMI) declined by more than 14 percent. At the intra-day level on this day of January 15, 2015, the drop was much worse with one euro exchanging temporarily against 0.86 CHF at the bottom of that day. Analysis of the exchange rate after January 15, 2015, is beyond the scope of this thesis. <sup>7</sup>

### 3.4.2 Uniqueness of the euro/Swiss franc target zone

We have seen in section 3.1 that Krugman's target zone model is based on two basic assumptions, marginal intervention and perfect credibility, and that it predicts a specific non-linear relationship between the (unobservable) fundamental  $f$  and the exchange rate  $s$  according to (3.10). Before testing the implied conditional volatility and probability density derived in section 3.3, we must make sure that Krugman's assumptions are sufficiently well fulfilled for the euro/Swiss franc target zone.

Qualitatively, the unique conditions of the EUR/CHF target zone are already visible when comparing the exchange rate fluctuations during the EUR/CHF target zone (Figure 2.1) with data from previous studies (for instance Figure 1 in [55], Figure 1b in [50] or any other empirical work cited in this chapter). Unlike in previous investigations, we notice for the EUR/CHF target zone a remarkable and sustained pressure that keeps the exchange rate close to the defended 1.20-barrier. To capture empirically this sustained strong pressure, in the next subsection, we will work with high resolution data of 10 second intervals. This enables us to sample also important intraday fluctuations close to the barrier. In particular, let us also highlight the exceptional conditions holding from April to September 2012 during which the exchange rate was practically indistinguishable from the barrier, as can be seen in Figure 2.1.

For a more quantitative analysis of the target zone environment, we rely on the work of Studer-Suter and Janssen [61] who have examined Krugman's assumptions extensively for the euro/Swiss

---

<sup>7</sup> However, let us point out that the active defense of the target zone led the SNB to accumulate foreign currency reserves of the order of the Swiss GDP. This very consideration is generally advanced to explain the decision of the SNB to abandon the cap on the Swiss franc's value against the euro on January 15, 2015, a week before the official announcement by the European Central Bank to launch an unprecedented quantitative easing program to buy an equivalent of 220% of the total net issuance planned by the Eurozone member states over the following 18 months. It would be interesting to examine whether an additional forward-rate target zone (as in [53]) could have helped making the target zone more sustainable.

franc from the inception of the target zone until January 2014. Different from Krugman's original approach, who describes the exchange rate as a function of exogenous velocity shocks, Studer-Suter and Janssen propose state variables for the exchange rate that come from a risk-based view of currency prices. This approach is particularly suitable for the Swiss franc as a safe haven currency. Hence, they replace Krugman's exogenous velocity  $v$  by  $-\kappa$  where a high  $\kappa$  indicates increased market risk. High market risk thus implies a lower  $s$ , which corresponds to a higher valuation for the Swiss franc against the euro. Studer-Suter and Janssen introduce a state variable, which unites all global and country-specific fundamentals and market sentiments that determine the EUR/CHF exchange rate. This state variable is given by the implied density function for the future value of the Swiss franc obtained from option prices. As it is derived from prices that emerge from trade in highly liquid markets, this state variable should incorporate all information that is relevant for the expected path of the EUR/CHF. Applying a unit-root test, they conclude that  $\kappa$  follows a random walk, which is in line with assumption (3.2). The assumption of marginal intervention is tested by considering the reserve holdings of the SNB. Noting that the reserves increase only when the exchange rate is exceptionally close to the lower limit, they conclude that marginal intervention most likely holds true. This contrasts with the experience from other exchange rate target zones where intra-marginal interventions have taken place frequently, thus rejecting the basic assumption of the Krugman.

In order to assess the credibility of the target zone, Studer-Suter and Janssen examine forward exchange rates as described in section 3.1. If markets believe that monetary authorities will always prevent the exchange rate from moving out of the target zone, forward exchange rates must never lie outside the band. For the EUR/CHF, this holds true for one month forward exchange rates, which suggest that, at the one-month horizon, investors never expected the Swiss franc to appreciate beyond CHF 1.20. However, forward rates at longer maturities noted well below this lower bound on several occasions. However, since the EUR/CHF exchange rate passes all tests on the one-month forward looking basis, Studer-Suter and Janssen conclude that perfect credibility is at least partially holding true and one can proceed with the analysis of Krugman's model.

Plotting the EUR/CHF exchange rate as a function of the EURO STOXX 50, which is the most important Blue-chip index of the Eurozone (a proxy for the perceived market risk  $\kappa$ ), yields widely scattered data points, resembling only roughly the non-linear relation (3.10). To improve on this state of affair, we will now test Krugman's model via the theoretical predictions of section

3.3, using our derived conditional volatility and probability density of the EUR/CHF exchange rate, which are directly measurable and do not require estimating market risks or a proxy for fundamentals. Specifically, we use the same data and measurements as presented in section 2.4.

### 3.4.3 Testing the smooth pasting condition

In the derivation of the conditional volatility (3.16), we have relied on the smooth pasting condition  $s(\underline{f}) = 0$ . However, our approach is flexible enough to relax this assumption, and test its validity directly from the data, instead. To this end, we start from the general form of the solution of the one-sided Krugman model

$$s(f) = f + Ae^{-\rho f} \quad (3.21)$$

but refrain from fixing  $A$  via smooth pasting condition. A second order Taylor expansion yields

$$s(f) = \underline{s} + \underbrace{(1 - A\rho \exp(-\rho \underline{f}))}_{\equiv \eta} (f - \underline{f}) + \frac{1}{2} \underbrace{(A\rho^2 \exp(-\rho \underline{f}))}_{\equiv \omega} (f - \underline{f})^2 + \mathcal{O}((f - \underline{f})^3). \quad (3.22)$$

Dropping third and higher order terms (see subsection 3.4.6), we can solve this quadratic equation for  $f - \underline{f}$ ,<sup>8</sup>

$$f - \underline{f} = \frac{\sqrt{\eta^2 + 2\omega(s - \underline{s})}}{\omega} - \frac{\eta}{\omega}. \quad (3.23)$$

In the same spirit as the derivation of (3.16), one calculates  $g(f) \approx \sigma A \rho^2 (f - \underline{f})$ , and, via (3.23),

$$g(s) = \frac{\rho^2 \sigma A}{\omega} \sqrt{\eta^2 + 2\omega(s - \underline{s})} - \frac{\rho^2 \sigma A \eta}{\omega}. \quad (3.24)$$

Defining  $D \equiv \rho^2 \sigma A / \omega$ , we have now derived the following expression for the conditional volatility:

$$g(s) = D \sqrt{\eta^2 + 2\omega(s - \underline{s})} - D\eta. \quad (3.25)$$

As expected, setting  $\eta = 0$  recovers (3.16), derived under the assumption that smooth pasting holds. We formulate the null hypothesis

$$H_0 : g(s) = D \sqrt{s - \underline{s}} \quad (3.26)$$

---

<sup>8</sup>The second solution,  $f - \underline{f} = -\frac{\sqrt{\eta^2 + 2\omega(s - \underline{s})}}{\omega} - \frac{\eta}{\omega}$ , is not economically sensible because  $f$  and  $s$  must be positively correlated.



and the alternative hypothesis

$$H_1 : g(s) = D\sqrt{\eta^2 + 2\omega(s - \underline{s})} - D\eta. \quad (3.27)$$

Clearly,  $H_0$  is nested in  $H_1$ . The likelihood ratio  $\Lambda$  is defined as

$$\Lambda = \frac{\text{maximum likelihood under } H_0}{\text{maximum likelihood under } H_1}. \quad (3.28)$$

Assuming that the error is independently and normally distributed with standard deviation  $\sigma_0$ , the likelihood function  $L_0$  under  $H_0$  reads

$$\begin{aligned} L_0 &= \prod_{\substack{\text{datapoints} \\ (s_i, g_i)}} \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{(g_i - g_0(s_i))^2}{2\sigma_0^2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^K \exp\left(-\frac{1}{2\sigma_0^2} \underbrace{\sum_{\substack{\text{datapoints} \\ (s_i, g_i)}} (g_i - g_0(s_i))^2}_{\equiv F_0}\right) \end{aligned} \quad (3.29)$$

where  $(s_i, g_i)$  are our measured datapoints plotted in Figure 2.3,  $i = 1, \dots, K$  with  $K = 100$ ,  $g_0(s)$  is the conditional volatility under  $H_0$  given by expression (3.26) and we have defined the cost function  $F_0$ . The standard deviation  $\sigma_0$  is given by  $\sqrt{\frac{1}{n-1}F_0}$ . Similarly, we define  $L_1, F_1$  and  $\sigma_1$  under hypothesis  $H_1$  with  $g_1(s)$  given by expression (3.27). According to Wilks' theorem, the distribution of the statistic  $-2\log(\Lambda)$  is asymptotically  $\chi^2$ -distributed with one degree of freedom. Numerically, we obtain

$$-2\log(\Lambda) = 0.8589 \quad (3.30)$$

which gives a  $p$ -value of  $0.36 > 0.05$ . Thus,  $H_0$  provides a description of the conditional volatility that is not improved significantly by  $H_1$ . Since the smooth pasting condition  $s(\underline{f}) = 0$  leading to  $H_0$  removes the need for one parameter ( $\eta$ ) it is preferred. Since the smooth pasting condition relies on a no-arbitrage argument, our result has a direct economic interpretation: it is likely that it was not possible to arbitrage the EUR/CHF target zone by betting on finite jumps in the exchange rate close to the lower barrier  $\underline{s}$ .

### 3.4.4 Testing the conditional volatility

The previous subsection tells us that the smooth pasting condition cannot be rejected empirically. We can thus move on to test prediction (3.16) for the conditional volatility. First, we apply a one parameter non-linear least squares fit to the parameter-free estimate depicted in Figure 2.3. The fit determines

$$\beta = (5.42 \pm 0.06) \cdot 10^{-3}, \quad (3.31)$$

showing that the standard error of the fit is small. Next, we apply another non-linear least squares fit but this time we fit two parameters  $\beta$  and  $\mu$  of the model

$$g(s) = \beta (s - \underline{s})^\mu. \quad (3.32)$$

The fit determines

$$\beta = (5.25 \pm 0.5) \cdot 10^{-3} \quad \text{and} \quad \mu = 0.49 \pm 0.03 \quad (3.33)$$

in good agreement with the theoretically predicted  $\mu = 1/2$ . To formalize this result, we formulate the null hypothesis

$$H_0 : \quad g(s) = \beta \sqrt{s - \underline{s}} \quad (3.34)$$

and the alternative hypothesis

$$H_1 : \quad g(s) = \beta (s - \underline{s})^\mu. \quad (3.35)$$

We calculate the likelihood ratio (3.28) and find

$$-2 \log(\Lambda) = 0.0128, \quad (3.36)$$

corresponding to a  $p$ -value of  $0.91 > 0.05$ , accepting  $H_0$ . We infer that the conditional volatility is indeed of the form (3.16) that we have extracted from Krugman's model up to second order.

### 3.4.5 Empirical estimation of the probability density of exchange rates

Using a kernel density estimator, we depict in Figure 3.3 the probability density from the EUR/CHF exchange rate data from September 6, 2011, to January 14, 2015. To test whether the data fits prediction (3.19), we have applied exactly the same procedure as above. A first one

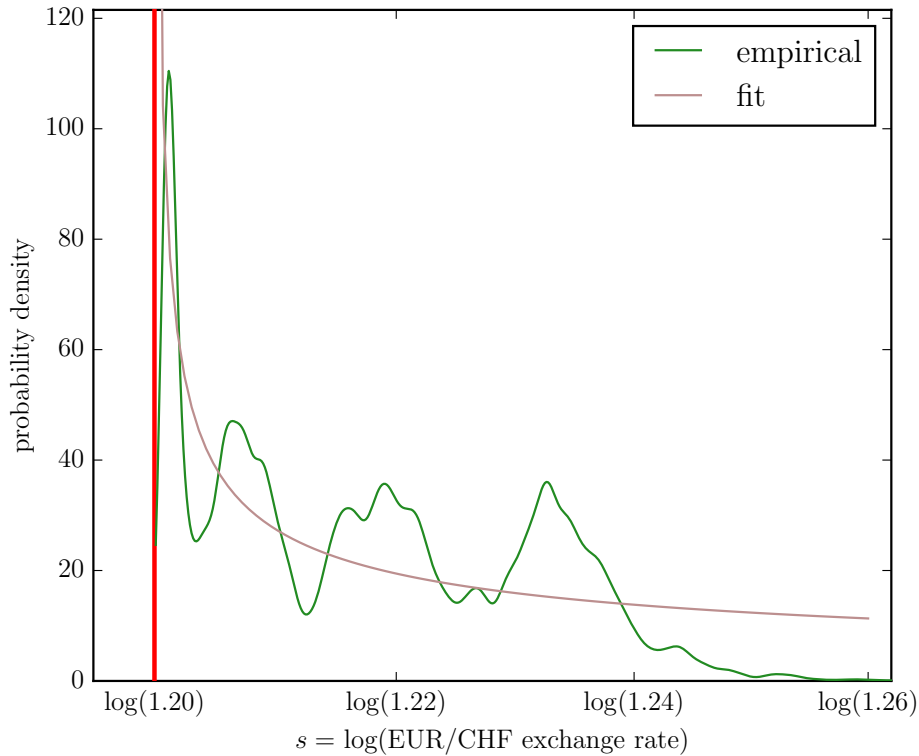


Figure 3.3: The green line shows the empirical distribution of EUR/CHF tick by tick data from September 6, 2011, to January 14, 2015, which, qualitatively, clearly exhibits the predicted probability density. The brown line denotes the one-parameter least squares fit to (3.19).

parameter non-linear least squares fit (shown in Figure 3.3) determines

$$C = 2.49 \pm 0.05. \quad (3.37)$$

Matters are slightly different when we make the two parameter fit for  $C$  and  $\nu$  to the model

$$p(s) = \frac{C}{(s - \underline{s})^\nu} \quad (3.38)$$

for which we obtain

$$C = 1.72 \pm 0.17 \quad \text{and} \quad \nu = 0.57 \pm 0.02. \quad (3.39)$$

We see that the estimated exponent  $\nu$  is still close to but not within one standard deviation from the theoretical prediction  $\nu = 1/2$ . This misfit is not a surprise but rather a necessary consequence of the singular behavior of (3.19). It is intuitively clear that the divergence in (3.19) can never be observed exactly but must be truncated due to finite size effects, leading to a maximum located at  $s^* \gtrsim \underline{s}$  instead of a true singularity at  $\underline{s}$ . If this was not the case,

i.e. if the SNB let the exchange rate approach too close to its imposed barrier, it may have to print arbitrary large amounts of Swiss francs, until an insufferable pain zone is reached, namely when the SNB balance sheet becomes of the order of or larger than the Swiss GDP, as a rough rule of thumb. In turn, this finite size effect induces an uncertainty about the location of the ‘true singularity’, which can then lead to large shifts in the estimated exponent  $\nu$ . It is then not surprising that the calculated log-likelihood ratio

$$-2 \log(\Lambda) = 15.48 \quad (3.40)$$

gives a  $p$ -value of order  $10^{-4} < 0.05$ , thus rejecting  $H_0$  in favor of  $H_1$  with exponent  $\nu = 0.57 \pm 0.02$ . Due to the necessarily induced shift between the theoretical singularity and the observed maximum, we conclude that the predicted  $\hookleftarrow$ -probability density can only be tested qualitatively (in the sense that the bulk of the probability mass lies close to the barrier) and not exactly at the quantitative level. We stress that, nevertheless, Figure 3.3 provides important support for Krugman's target zone model. Apart from the qualitative agreement with the predicted  $\hookleftarrow$ -shaped density, it shows that (modulo oscillations) the density increases as a function of the distance to the barrier with a peak very close to  $\underline{s}$  followed by a rapid decay. Moreover, the exponent  $\nu$  is only about 14% away from the theoretical prediction. Overall, this provides further evidence for the assumption of marginal intervention and sets the EUR/CHF target zone apart from many previous studies in which the bulk of the density has been observed further away from the barrier.

In order to relate the estimated value of parameter  $C$  (3.20) with (3.19) to economic observables, we must have an estimate of the fundamental density  $\tilde{p}$ . Because the fundamental variable is (asymptotically) uniformly distributed, it holds that

$$\tilde{p} = \frac{1}{\text{range of fundamental}}.$$

Now, what is the range of fundamental values? Under the free float, we could observe exchange rates as low as 0.85 CHF per euro. On the other hand, we can assume that the fundamental was not above 1.26 Swiss francs per euro (the maximal observed exchange rate value during the target zone policy). Therefore, we estimate

$$\tilde{p} = \frac{1}{\log(1.26) - \log(0.85)} \quad (3.41)$$

and we assume a relatively large standard error of  $\frac{1}{2}(\log(1.26) - \log(0.85))$  to account for possible deviations that could not be observed. The parameter  $\rho = \sqrt{2/\gamma\sigma^2}$  then derives from  $\gamma$  and  $\sigma$  with basic error propagation, to obtain

$$\rho = 0.61 \pm 0.21. \quad (3.42)$$

Having estimated both  $\tilde{p}$  and  $\rho$ , we finally derive the value of the normalisation constant

$$C = 2.31 \pm 0.44, \quad (3.43)$$

which is good agreement with the fitted parameter (3.37), showing the overall consistency of our estimations of the key predictions (3.16) and (3.19) of Krugman's model.

We finish this section by pointing out that both Figures 2.3 and 3.3 exhibit rather well-defined oscillations decorating the square-root and inverse-square root laws (3.16) and (3.19). The local maxima observed in Figure 3.3 correspond to local concentrations of exchange rate values, and we find that they are located at

$$\text{EUR/CHF} = 1.2020, 1.2064, 1.2190, 1.2326, 1.2437.$$

Converting to the ratio CHF/EUR, these maxima occur at

$$\text{CHF/EUR} = 0.8319, 0.8288, 0.8203, 0.8112, 0.8040,$$

where are within 20% of CHF/EUR = 0.83, 0.82, 0.81, 0.80, suggesting anchoring by traders on discrete values.

### 3.4.6 Justification of the Taylor approximations

Here, we show that the second order Taylor approximations that we have used in section 3.3 are justified for the EUR/CHF target zone. The next order contribution to the Taylor expansion of  $s(f)$ , (3.12), is given by  $\frac{1}{6}\rho^2(f - \underline{f})^3$ . We have argued in section 3.4.5 that, although a priori modeled as a free random walk, the fundamental  $f$  fluctuates only between  $\log(1.26)$  and a lower bound of approximately  $\log(0.8)$ , so that  $f - \underline{f} < 1/2$  held at all times during the target zone arrangement. We have also shown that (3.42) =  $\rho < 1$ . It derives immediately that the higher

order corrections are negligible by several orders of magnitude.

For justification of the approximations in  $g(s)$ , we compare the first to the second order term, given in (3.15). Using the estimate (3.42), it is easy to check that, in the worst case scenario where  $\rho = 0.61 + 0.21$  and as far away from the barrier as  $s = \log(1.26)$ , the first order correction is still more significant by a factor larger than 7. This significance increases rapidly the closer to the barrier and the smaller the actual value of  $\rho$ . Also, since the ratio of the first to the second order depends only on the square root of the parameters, the approximation is robust with respect to variations of  $\sigma$  and  $\gamma$  over several orders of magnitude.

One could point out that it is logically inappropriate to justify the Taylor approximation by assuming  $\rho < 1$  since the numerical value of  $\rho$  itself was calculated from a fit for  $\beta$  that, in turn, is based on the assumption that the Taylor approximation is justified. Assume that we were wrong to neglect the next order correction (3.15) (which comes with a minus sign!). Then, we would have estimated a  $\beta$  that is smaller than it really is,  $\beta_{\text{real}} > \beta_{\text{fitted}}$ . This, in turn, means that the implied  $\sigma$  would be larger than we have estimated and the real  $\rho \sim 1/\sqrt{\sigma}$  would be even smaller than what we have estimated. In conclusion, the Taylor approximation (3.12), for the EUR/CHF target zone, is completely justified a posteriori.

More formally, we can appeal to the standard formal approach in applied mathematics that we now sketch. Consider a problem  $P$ , whose set of solutions is  $\{S\}$ . Suppose that we can approximate  $P$  by  $P_a$  and this approximation is valid only for those solutions in  $\{S\}$  that belong to a certain domain  $V \subset \{S\}$ . Let us consider the set of all solutions  $\{S_a\}$  of problem  $P_a$ . If these solutions  $\{S_a\}$  belong to the domain  $V$ , then the solutions  $\{S_a\}$  approximate the true solutions  $\{S\}$  of the initial full problem  $P$ .

In order to show that the first order correction in (3.19) is justified, we must show that the contribution of the third order expansion of  $f(s)$  is negligible in the derivative  $df/ds$  as long as  $s$  is close to  $\underline{s}$ . Expanding  $s(f)$  in (3.12) to third order yields

$$s = \underline{s} + \frac{1}{2}\rho(f - \underline{f})^2 - \frac{1}{6}\rho^2(f - \underline{f})^3 \quad (3.44)$$

which is a cubic equation in  $f - \underline{f}$  and hence tedious. Instead, we use the inversion theorem

$$\frac{df}{ds} = \frac{1}{\frac{ds}{df}} \quad (3.45)$$

and show that higher order correction terms to  $\frac{ds}{df}$  are negligible. From (3.44), we calculate

$$\frac{ds}{df} = \rho(f - \underline{f}) - \frac{1}{2}\rho^2(f - \underline{f})^2 \quad (3.46)$$

and since both  $f - \underline{f}$  and  $\rho$  are smaller than 1, it is evident that the second order term in  $ds/df$  is negligible by several orders of magnitude. This translates into

$$\frac{df}{ds} = \frac{1}{\frac{ds}{df}} = \frac{1}{\left(\frac{ds}{df}\right)_{\text{first order}} + \left(\frac{ds}{df}\right)_{\text{second order}} + \dots} \approx \frac{1}{\left(\frac{ds}{df}\right)_{\text{first order}}} \approx \left(\frac{df}{ds}\right)_{\text{first order}} \quad (3.47)$$

and justifies the last first order expansion of section 3.3.

### 3.5 Conclusions

In this chapter, we have reexamined the classic target zone model by Krugman [40] which, based on a minimal amount of assumptions, describes the dynamics of an exchange rate confined to lie within a certain range of values. Despite its wide acceptance from a theoretical point of view, empirical support has been sparse and empirical tests have most often led to rejections of the model. Criticism was based on two arguments. First, there was no convincing evidence that the assumptions of perfect credibility and marginal interventions hold true. Second, it turned out to be difficult to capture the subtle non-linear relationship between the exchange rate and the underlying unobservable fundamental value.

We have shown that, compared to previous studies, the EUR/CHF target zone is unique in the sense that the ‘pressure’ against the 1.20 CHF per euro barrier was exceptionally high and, at the same time, the defence by the central bank extremely rigorous. With help from the thorough analysis by Studer-Suter and Janssen [61], we could formalise this observation and argued that the assumptions of Krugman’s model apply sufficiently well so that the model can be tested on EUR/CHF exchange rate data, thus refuting the first objection to the validity of the model.

We have addressed the second point of criticism, namely the lack of evidence of the predicted honey-moon effect, by proposing a more direct test of the model. Showing that, close to the barrier, the dependence of Krugman’s exchange rate as a function of the fundamental can be locally inverted, we have derived analytical expressions both for the conditional volatility and for the probability density function of the exchange rate as a function of the distance to the barrier.

We have also presented a statistical test for the smooth pasting condition, associated with the no-arbitrage condition in Krugman's model. These predictions were subsequently tested on high resolution EUR/CHF exchange rate data from September 2011 to January 2015. Extracting the conditional volatility from a parameter-free estimation and applying a likelihood ratio test, we have provided, to the best of our knowledge for the first time, direct empirical and nonparametric evidence that the bounded EUR/CHF exchange rate is well described by Krugman's target zone model. From the calibrated parameters, we could further derive a large implied volatility for the fundamental variable during the time when the exchange rate target was imposed, which arguably reflects the large perceived uncertainties of these times. Estimating the empirical distribution of exchange rates, we have provided qualitative and semi-quantitative evidence that the density is of the predicted  $\hookleftarrow$ -shaped form. In conclusion, we have shown that Krugman's target zone model is supported by empirical data after all. However, and this is what distinguishes our findings from previous studies, this holds only under extreme and sustained pressure (associated with the European crisis) that pushed continuously the exchange rate very close to the boundary of the target zone.



## Chapter 4

# An option representation of currency target zones †

In chapter 2, we have discussed target zones from the perspective of constrained physical random walks. In chapter 3, we have seen how target zones are more appropriately described with economic models that explicitly take into account the expected change in the underlying exchange rate. In this chapter, we make use of this insight by describing target zones via options. Indeed, options are forward-looking derivatives explicitly designed to take into account future expectations. In the next section, we will then map the option representation to the classical Krugman model of the previous chapter.

During the era of the European Exchange Rate Mechanism (ERM) established to prepare the transition of European currencies towards the euro as a single currency, target zones have emerged as a popular macroeconomic tool of central banks. In a target zone arrangement, the central bank pledges to keep its domestic currency within a pre-specified band with respect to some other currency. The Hong-Kong Monetary Authority (HKMA) and the Swiss National Bank (SNB) are two prominent recent examples of target zones. <sup>1</sup>

The importance of understanding the far reaching implications of such target zones has sparked a myriad of research over the last two decades. At its origin is the model by Krugman [40], discussed in the previous chapter. To account for deviations from the strict assumptions in the

---

†This chapter is based on reference [11].

<sup>1</sup> Since 2005, the HKMA has been regulating its currency such as to keep it between 7.75 and 7.85 Hong-Kong dollar per US dollar (Figure 4.3). Between September 6, 2011 and January 15, 2015, the SNB defended its continuously strengthening Swiss-franc by enforcing a minimum exchange rate of 1.20 Swiss-francs per euro. Since the exchange rate is only bounded from one side, this represents a special case of a single-sided target zone. See chapter 2 and 3.

Krugman model, several extensions have been developed, most notably to include intramarginal interventions [46], imperfect credibility [44], soft target boundaries [54], breaching probability [73], forward-looking target zones [53], or market inefficiency [74]. See ref. [75] for an in-depth overview.

In this chapter, we present a different approach that allows to view the target zone from a completely different angle. Concretely, we show how target zones can be seen as an infinite series of ‘mirrored’ American put- and call- options. Finding an alternative mathematical formulation has the advantage that certain aspects may appear difficult in one, but rather simple in the other formulation. For instance, while it can be difficult to assess the credibility of a target zone in its original formulation, we will see how it has a very intuitive interpretation within option pricing.

The idea of an analogy between target zones and options is not new. Already Krugman [40] pointed out a qualitative connection between a two-sided target zone and option pricing. Much later, this idea was picked up and addressed more quantitatively by Nazsodi [76]. However, as we shall see in section 5.2, her arguments are incomplete. For single-sided target zones, the representation of target zones in terms of options is extremely simplified, as will become clear in section 5.2. This special case has been discussed in the context of the recent euro-Swiss franc target zone. Hanke et al. [77] make use of the relation to option pricing to estimate the fundamental, unobserved exchange rate that would have prevailed in absence of a target zone. Jermann [78] presents a discrete model that takes into account the perceived probability of an ending target zone policy. Hanke et al. [79] use a latent likelihood approach to estimate the fundamental and expected remaining lifetime of the target zone. We will come back to their method in section 4.4.

Instead of relying on special cases, we contribute to the existing literature by describing exactly, for the first time, the two-sided target zone model using compound options. Due to so-called mirror-effects, this mapping requires an infinite sequence of nested option prices. Our exact solution is conceptually more involved than previous approximate approaches, but it has many conceptual advantages. To give a concrete example, it allows for a natural estimate of target zone credibility. Furthermore, this mapping can be used to model different target zone arrangements conveniently.

The remainder of this chapter is structured as follows. Section 5.2 describes two-sided target zones in terms of option prices, which constitutes the core result of this chapter. This allows us to derive an alternative way of estimating the underlying fundamental value, which is explained

theoretically in section 4.4 and applied to the US dollar/Hong-Kong dollar exchange rate in section 4.5. Section 4.6 concludes.

## 4.1 Target zones as infinite sequences of compound options

In this section, we derive a mapping between target zones and option prices. To understand the relation between target zones and option prices, it is helpful to think first of a single-sided target zone, implemented only on the strong side of the domestic currency. Assume holding one unit of foreign currency, which is worth  $S$  units of domestic currency.<sup>2</sup> If there is no target zone in place, this unit of foreign currency is worth as much as its fundamental value, determined by the efficient market, i.e.  $S = F$ . Now, the central bank steps in and announces to defend a lower boundary at  $\underline{S}$ . This is formally equivalent to saying the central bank offers the holder of the foreign currency an additional put option for free, i.e.

$$S = F + \mathcal{P}(F, \underline{S}) \quad (4.1)$$

where  $\mathcal{P}(F, \underline{S})$  denotes an American put option with strike price  $\underline{S}$  and underlying  $F$ . The option is American since it can be exercised any time. Within the classic model of Krugman [40], it is assumed that the target zone policy is perfectly credible and held ‘forever’, i.e. maturity  $\tau$  is infinite. Such options are called perpetual. In the option pricing formalism, there is no need to stick to this assumption;  $\tau$  can take any positive value. In section 4.4, we show how  $\tau$  can be estimated from the data. This lets us interpret  $\tau$  as a proxy of the perceived target zone credibility. To simplify notation, the dependence on  $\tau$  is usually not stated explicitly in the remainder of this chapter.

Arguments similar to the ones leading to (4.1) apply for a single-sided target zone on the weak side. A central bank announcing a weak-side target zone restricts the value of the asset to a maximum value of  $\bar{S}$ . In the language of option pricing, the holder of the asset is now short an American call option  $\mathcal{C}$  with strike price  $\bar{S}$ , i.e.  $S = F - \mathcal{C}(F, \bar{S})$ . Transitioning to the general two-sided target zone, the naive idea that comes to mind is

$$S = F + \mathcal{P}(F, \underline{S}) - \mathcal{C}(F, \bar{S}), \quad (4.2)$$

---

<sup>2</sup> Note that in the previous chapters, we have worked with the logarithms of the actual quantities,  $s = \log(S)$ , and so forth.

which is, however, not correct. To understand where this reasoning fails, think about what happens if one of the options, say the put, is exercised. Exercising means the one unit of foreign currency is sold in exchange for  $\underline{S}$  units of domestic currency. In that case, one does not only get rid of the unit of foreign currency, but one also gets rid of the associated options that came along with the target zone arrangement, i.e. the call one was short with. The reasoning for the short call option is analogous. To first order, we can therefore write

$$S = F + \mathcal{P}(F - \mathcal{C}(F, \bar{S}), \underline{S}) - \mathcal{C}(F + \mathcal{P}(F, \underline{S}), \bar{S}). \quad (4.3)$$

The argument does not stop here, but has to be iterated.<sup>3</sup> Consider the following analogy to mirrors. Placing one mirror results in a single reflection of the object. Placing the objects between two mirrors produces an infinite number of reflections, namely the reflection, the reflection of the reflection, and so on.<sup>4</sup> And in the same manner, exercising the put does not only lead to an exchange of the foreign currency unit for the fundamental minus the shorted call of the fundamental, but that call is itself is again a function of the put, and so on. The dynamics of the exchange rate within a target zone is thus given by the dynamics of the value of the following artificial portfolio of nested compound options

$$S = F + \mathcal{P}(F - \mathcal{C}(F + \mathcal{P}(F - \mathcal{C}(\dots)))) - \mathcal{C}(F + \mathcal{P}(F - \mathcal{C}(F + \mathcal{P}(\dots))), \quad (4.4)$$

where we have dropped the strike prices for notational simplicity. We stress that the options in (4.4) are iteratively written on the unobservable fundamental  $F$  and not on the exchange  $S$ . As such, these option prices cannot be observed but are merely a useful mathematical abstraction.

5

We define the  $n$ -th order approximation of relation (4.4) as the sequence

$$S_n(F) = \begin{cases} f + \mathcal{P}_n(F - \mathcal{C}_{n-1}(F \pm \dots + \mathcal{P}_1(F))) - \mathcal{C}_n(F + \mathcal{P}_{n-1}(F \mp \dots - \mathcal{C}_1(F))), & \text{if } n \text{ is odd,} \\ f + \mathcal{P}_n(F - \mathcal{C}_{n-1}(F \pm \dots - \mathcal{C}_1(F))) - \mathcal{C}_n(F + \mathcal{P}_{n-1}(F \mp \dots + \mathcal{P}_1(F))), & \text{if } n \text{ is even.} \end{cases} \quad (4.5)$$

<sup>3</sup> Similar arguments were brought forward by Naszodi [76]. However, she incorrectly stops at (4.3). Taking into consideration higher order correction terms can be of utmost importance for narrow target zones, as is seen in Figure 4.1.

<sup>4</sup> Similar ideas have been applied in electrostatics and random walk theory, in what is usually referred to as ‘the method of images’ [25, 80].

<sup>5</sup> The pricing of actually tradable options for the exchange rate  $S$  bound in a credible target zone has been investigated in ref. [81].

We also define  $S_0 = F$  as the special case of a freely floating exchange rate. Assume the underlying  $F$  (under the risk neutral measure) follows a stochastic process of the form

$$dF = (r - \delta) F dt + \sigma(F) dW_t \quad (4.6)$$

for some volatility function  $\sigma$ .<sup>6</sup> The parameters  $r$  and  $d$  denote the domestic and foreign risk free rate, respectively. Formally, relation (4.5) constitutes the main result of this chapter. In the following, we discuss its convergence to the actual exchange rate  $S$  for  $n \rightarrow \infty$ , its numerical implementation and its calibration to exchange rate data.

## 4.2 Iterative implementation of mirrored option prices

The innermost functions in (4.5),  $\mathcal{C}_1, \mathcal{P}_1$ , are American options with underlying process (4.6) and strike prices  $\underline{S}$  and  $\bar{S}$ , respectively. At order  $k \in \{2, \dots, n\}$ , the option price  $\mathcal{C}_k, \mathcal{P}_k$  are no more written on the underlying (4.6) but on a stochastic process that depends recursively on  $F$  and lower order option prices. The iteration starts from

$$\mathcal{P}_1(F) \equiv \mathcal{P}(F, \sigma, \underline{S}, \tau) \quad (4.7a)$$

$$\mathcal{C}_1(F) \equiv \mathcal{C}(F, \sigma, \bar{S}, \tau) \quad (4.7b)$$

where  $\mathcal{P}(F, \sigma, \underline{S}, \tau)$  denotes the price of an American put option written on  $F$  with maturity  $\tau$ , strike price  $\underline{S}$  and underlying volatility function  $\sigma$ . Similarly,  $\mathcal{C}(F, \sigma, \bar{S}, \tau)$  is the price of an American call option with strike price  $\bar{S}$ . In the following, we drop the parameters  $\underline{S}, \bar{S}$  and  $\tau$  for notational simplicity and since they do not change. Here,  $\sigma$  is a function, and not a constant like for geometric Brownian motion.<sup>7</sup>

Pricing  $\mathcal{P}(F, \sigma)$  and  $\mathcal{C}(F, \sigma)$  for general functions  $\sigma$  is known as local volatility modelling. For all practical cases, prices have to be determined numerically. We have therefore implemented a binomial tree model [82] in C++.<sup>8</sup> Starting from (4.7) and equipped with a numerical pricer of

<sup>6</sup> The special case of geometric Brownian motion is recovered for  $\sigma(F) = \alpha \cdot F$  with  $\alpha > 0$  some constant.

<sup>7</sup> Assuming that the underlying follows a geometric Brownian motion represents the special case where  $\sigma(F) = \alpha \cdot F$  for some constant  $\alpha > 0$ .

<sup>8</sup> Since the local volatility binomial tree can have instabilities, various extensions have been proposed [83]. For our purposes, the binomial tree was sufficient.

$\mathcal{P}(F, \sigma)$  and  $\mathcal{C}(F, \sigma)$ , one can then calculate relation (4.5) iteratively according to

$$\mathcal{C}_{k+1}(F) = \mathcal{C}(F + \mathcal{P}_k(F), \sigma_{k+1}^{\mathcal{C}}) \quad (4.8a)$$

$$\mathcal{P}_{k+1}(F) = \mathcal{P}(F - \mathcal{C}_k(F), \sigma_{k+1}^{\mathcal{P}}) \quad (4.8b)$$

where the volatility functions  $\sigma_{k+1}^{\mathcal{C}}, \sigma_{k+1}^{\mathcal{P}}$  are themselves determined recursively,

$$\sigma_{k+1}^{\mathcal{C}}(\cdot) = \sigma(\cdot) (1 + \mathcal{P}'_k(\cdot)) \quad (4.8c)$$

$$\sigma_{k+1}^{\mathcal{P}}(\cdot) = \sigma(\cdot) (1 - \mathcal{C}'_k(\cdot)), \quad (4.8d)$$

with the prime denoting the derivative. To see this, recall that  $F$  obeys (4.6). The option  $\mathcal{C}_{k+1}$  is written on the underlying  $F + \mathcal{P}_k(F)$ . The corresponding stochastic process is then

$$d(F + \mathcal{P}_k(F)) = dF + d\mathcal{P}_k(F) = (\dots) dt + \left( \sigma(F) + \sigma(F) \frac{\partial \mathcal{P}_k}{\partial F}(F) \right) dW_t \quad (4.9)$$

according to Itô's lemma. A similar expression holds for the process of  $F - \mathcal{C}_k(F)$  that underlies the put option  $\mathcal{P}_{k+1}$ . We omit stating here explicitly the drift term since it will be replaced by the interest rate differential when pricing under the risk neutral measure. This finishes the iterative description of procedure (4.5).

When implementing system (4.8), a recursive implementation appears reasonable. However, since each function calls itself two functions, the number of function evaluations scales exponentially in  $n$ . Considering that each option call spans a binomial tree, and the derivative needs to be approximated numerically, this becomes computationally infeasible. Even with an efficient C++ implementation, we cannot go higher than  $n = 4$  within a reasonable amount of time. A forward implementation is thus required. One caveat of the forward implementation is that the options and volatility functions recursively depend on all previous iterations. To overcome this problem, we approximate in each step the options through spline interpolations. This is computationally inexpensive and the options can be approximated with arbitrary precision. The advantage of this approach is that the options and volatilities at iteration step  $k + 1$  now only depend on the volatility at step  $k$ . Our implementation has been written in Python with an interface to C++ for the binomial option pricing.

### 4.3 Proof of convergence

To proof that  $S_n$  from relation (4.5) converges to the actual exchange rate  $S$ , we define a dynamical system through

$$x_{n+1} = f + \mathcal{P}(y_n) \quad (4.10a)$$

$$y_{n+1} = f - \mathcal{C}(x_n) \quad (4.10b)$$

with  $x_0, y_0 = F$ . It is straightforward to see that

$$S_n = x_n + y_n - F \quad (4.11)$$

is equivalent to the  $n$ -th order iteration of system (4.5).<sup>9</sup> The fixed point obeys

$$x^* = f + \mathcal{P}(y^*) \quad (4.12a)$$

$$y^* = f - \mathcal{C}(x^*). \quad (4.12b)$$

Consider the neighborhood of the fixed-point  $(x^*, y^*)$ . First order Taylor approximations give

$$x_{n+1} = f + \mathcal{P}(y^*) + \left. \frac{d\mathcal{P}}{dy} \right|_{y=y^*} (y_n - y^*) + \mathcal{O}(y_n - y^*)^2 \stackrel{(4.12)}{\approx} x^* + \mathcal{P}'(y^*)(y_n - y^*) \quad (4.13a)$$

$$y_{n+1} = f - \mathcal{C}(x^*) - \left. \frac{d\mathcal{C}}{dx} \right|_{x=x^*} (x_n - x^*) + \mathcal{O}(x_n - x^*)^2 \stackrel{(4.12)}{\approx} y^* - \mathcal{C}'(y^*)(x_n - x^*). \quad (4.13b)$$

Defining  $\delta x_n = x_n - x^*$  and  $\delta y_n = y_n - y^*$ , system (4.13) can be written in matrix form and iterated,

$$\begin{pmatrix} \delta x_n \\ \delta y_n \end{pmatrix} = \begin{pmatrix} 0 & \mathcal{P}'(y^*) \\ -\mathcal{C}'(x^*) & 0 \end{pmatrix} \begin{pmatrix} \delta x_{n-1} \\ \delta y_{n-1} \end{pmatrix} = \dots = \begin{pmatrix} 0 & \mathcal{P}'(y^*) \\ -\mathcal{C}'(x^*) & 0 \end{pmatrix}^n \begin{pmatrix} \delta x_0 \\ \delta y_0 \end{pmatrix}. \quad (4.14)$$

The vector  $(\delta x_0, \delta y_0)$  is interpreted as an initial slight deviation from the fixed point  $(x^*, y^*)$ .

For options priced under geometric Brownian motion, it is well known that put/call option prices are monotonically decreasing/increasing in  $F$ , and that

$$0 \geq \mathcal{P}'(F) \geq -1 \quad \text{and} \quad 0 \leq \mathcal{C}'(F) \leq 1 \quad \forall f. \quad (4.15)$$

---

<sup>9</sup> To simplify notation, we omit here the fact that the volatility functions  $\sigma_k^{\mathcal{P}, \mathcal{C}}$  also have to be iterated. This has no influence on our results. Just like in the previous section, we do not write strike prices and maturities explicitly since they are held constant,  $\mathcal{P}(F) \equiv \mathcal{P}(F, \sigma, \tau, \underline{S})$ ,  $\mathcal{C}(F) \equiv \mathcal{C}(F, \sigma, \tau, \bar{S})$ .

Equality holds only for zero volatility, at maturity or asymptotically when infinitely deep in- or out- the money. These cases are not relevant for our problem and so we can assume strict inequality. It can be shown that (4.15) extends to more general stochastic processes of the underlying [84], in particular the continuous local volatility models we are interested in. The eigenvalues of the matrix in equation (4.14) are thus found to be real-valued and distinct,  $\lambda_{\pm} = \pm\sqrt{-\mathcal{P}(y^*)\mathcal{C}(x^*)}$ . Diagonalizing the matrix we then express system (4.14) as

$$\begin{pmatrix} \delta x_n \\ \delta y_n \end{pmatrix} = \Lambda \begin{pmatrix} \lambda_+^n & 0 \\ 0 & \lambda_-^n \end{pmatrix} \Lambda^{-1} \begin{pmatrix} \delta x_0 \\ \delta y_0 \end{pmatrix}. \quad (4.16)$$

with  $\Lambda$  the transformation matrix. Since both eigenvalues are between 0 and 1,  $\lim_{n \rightarrow \infty} \lambda_{\pm}^n = 0$  and the system is therefore (locally) stable, i.e.  $S_n \rightarrow S$  as claimed. Counterintuitively, we find that the convergence of  $S_n$  to  $S_{\infty}$  is very slow and the use of high order iterations of (4.5) is crucial. For instance, to approximate a target zone such as the one employed for the Hong Kong dollar, approximations up to at least order 50 are required. It is also interesting to see that the second order correction (4.3) is generally worse than the naive first order idea (4.2), indicating that that the convergence  $S_n(F) \rightarrow S$  is not monotonic. Examples are depicted in Figure 4.1.

While the numerical approximation of equation (4.4), approximated by increasing iteration orders of (4.5) is technically challenging, it is conceptually very convenient. For instance, option prices come with a maturity parameter, which has a natural interpretation in terms of target zone credibility. The more credible the target zone is perceived by the market, the longer the implied maturity. As is discussed in subsequent sections, this maturity can be estimated from the data using maximum likelihood to assess the target zone credibility. The classical assumption of perfect credibility as assumed by Krugman [40] is formally recovered by the maturity going to infinity. But also other target zone models have a natural interpretation in terms of exotic options. For instance, a forward-looking target zone [53] is just given by (4.4) with the options written on future fundamental prices. Market inefficiencies [74] may alternatively be modeled with option pricing methods for inefficient markets. To give another example, consider the idea of a soft target zone, i.e. a target zone where the boundaries are not strictly enforced, but are enforced on average [54]. Small fluctuations around the boundaries are allowed, e.g. to discourage arbitrageurs. We claim that such a target zone can be represented by barrier options that allow to exercise within a specific price band. Finally, many techniques are known to price options if the fluctuations in the underlying are non-Gaussian. Such results can directly be translated



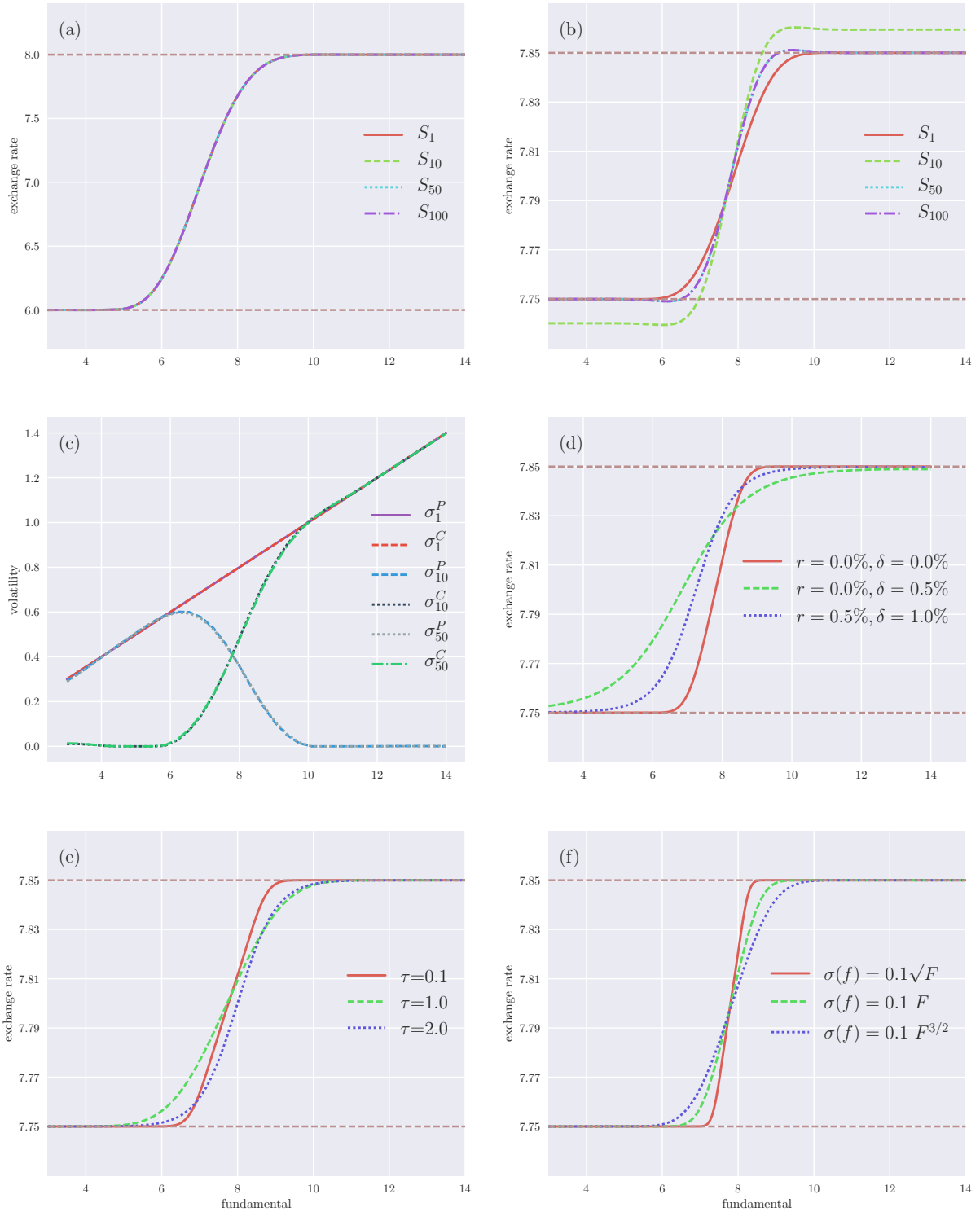


Figure 4.1: Different examples of the solutions to system (4.5) are shown. Unless stated otherwise, all figures show  $S_{100}(F)$  with the following parameters:  $\sigma(F) = 0.1F$ ,  $r = 0\%$ ,  $\delta = 0\%$ ,  $\tau = 1$  year,  $\underline{S} = 7.75$ ,  $\bar{S} = 7.85$ . These boundaries represent the Hong-Kong dollar/US dollar target zone discussed in section 7.4. Figure (a) is an example of  $S_n$  in a wide target zone, where  $\underline{S} = 6$ ,  $\bar{S} = 8$ . Higher iteration orders do not play an important role in this case as the two boundaries are essentially independent of each other. The strike prices are sufficiently far apart so that the put is deep out of the money at  $\bar{S}$  and the call is deep out of the money at  $\underline{S}$ . In contrast, Figure (b) highlights the importance of higher iteration orders in narrow (but realistic!) target zones. Convergence is relatively slow, and only for  $n = 50$  does the exchange rate approximation  $S_n$  join the target zone boundaries. Figure (c) complements Figure (b) by showing the respective volatility functions of the underlying nested process. Higher order volatilities deviate significantly from the geometric Brownian motion  $\sigma_1^{P/C}$  of the true underlying fundamental  $F$ . Figure (d) exemplifies the influence of domestic and foreign interest rates  $r$  and  $\delta$ . The difference in interest rate controls the skewness of the relation between exchange rate and fundamental. Figure (e) visualizes the effect of maturities. The larger the maturity, the more is the option worth and the flatter is the  $S$  curve. A similar observation is made in Figure (f) for different volatilities, since larger volatility typically leads to larger option prices.

to target zones through the mapping we have provided. Detailed comparisons of the different approaches, explicit mappings and more applications of this kind are part of future research.

#### 4.4 Estimating maturity and fundamental value

In the previous section, we have derived a relationship between the exchange rate  $S$ , and the fundamental value  $F$  that would prevail in absence of the target zone. A large discrepancy between exchange rate and fundamental implies a high risk for enterprises involved in business across borders. In case the target zone policy was suddenly abandoned, the exchange rate immediately jumps towards its fundamental value, sometimes with transient overshooting. Such an unexpected jump of the exchange rate value severely impacts businesses or even entire economies.

<sup>10</sup> Beyond the fundamental value, from the view of the central bank, knowledge of the target zone credibility is also crucial. As was noted by Krugman [40], a credibly maintained target zone leads to the market endogenizing the target zone, without the central bank having to intervene actively (So-called honeymoon effect). In contrast, if the central bank has to prove its commitment, significant and potentially costly interaction on the foreign exchange market is required. Both the fundamental value and the credibility are not directly observable. Using maximum likelihood, we show now how to estimate this fundamental value and the target zone credibility.

Since options are priced under the risk-neutral measure, it is fairly general to assume that the fundamental follows a process (4.6) with  $\sigma(F; \theta)$  a function that depends on a parameter  $\theta$  to be estimated. The parameter  $\theta$  can be a vector and represent a number of parameters. We write  $S = S(F; \theta, \tau)$  to denote the functional relation between exchange rate  $S$  and fundamental  $F$  with  $\tau$  being the implied maturity. To simplify notation, we absorb the  $\tau$  dependence into the vectorial  $\theta$  parameter and write  $S = S(F) = S(F; \theta)$ . It is not hard to see from Figure 4.1 that this functional relation is always invertible, so that we can also write  $F(S)$ . <sup>11</sup>

Denote by  $p_F(\Delta F | F_0, \Delta t, \theta)$  the probability density of the fundamental moving a step  $\Delta F$  within a timespan  $\Delta t$ , conditional on its current value  $F_0$ . This transition probability is

---

<sup>10</sup> An illustrative example of this case is the single-sided euro/Swiss-franc target zone that prevailed from September 6, 2011 until January 15, 2015. Without any prior announcement, and to the surprise of many, the Swiss Central Bank abandoned its currency defense on January, 15, 2015, leading to an extremely fast intraday exchange rate drop from 1.2 to below 0.85, which finally settled at 1.05 Swiss francs per euro at the end of the day. The Swiss Market Index (SMI) declined by more than 14 percent on January 15, 2015, with much larger transient intraday drops and high volatility.

<sup>11</sup> More formally, any positive/negative change in  $S$  must necessarily translate into a positive/negative change in  $F$  and vice versa. The function is then injective and trivially surjective on its domain of definition. This makes the function uniquely invertible.

determined as the solution to the Kolmogorov-equation<sup>12</sup> relating  $p_F$  to its stochastic counterpart (4.6). Through  $S(F; \theta)$ , we then derive a transition probability for the observable exchange rate,

$$p_S(\Delta s \mid S_0, \Delta t, \theta) = \left| \frac{dF}{dS} \right| p_F(\Delta F \mid F_0, \Delta t, \theta), \quad (4.17)$$

where the right-hand side of (4.17) is understood as a function of  $S$  through  $F = F(S; \theta)$ . Assume that we have observed some  $k + 1$  exchange rate datapoints  $\{S_i(t_i)\}_{i=1}^{k+1}$  at times  $t_i$ . We extract  $k$  exchange rate fluctuations  $\Delta S_i = S_{i+1} - S_i$  with time steps  $\Delta t_i$ . This allows us to determine the estimate  $\hat{\theta}$  of the parameter  $\theta$  as the argument that maximizes (the logarithm of) the likelihood function

$$\mathcal{L}(\theta) = \prod_{i=1}^k p_S(\Delta S_i \mid S_i, \Delta t_i, \theta). \quad (4.18)$$

The estimated fundamental value  $\hat{F}_i$  at time  $t_i$  is finally determined through  $\hat{F}_i = F_n(S_i; \hat{\theta})$  where  $F_n$  is the inverse of  $S_n$  given by (4.5).

A similar approach has been conducted by Hanke et al. [79], who studied the single-sided euro-Swiss franc target zone that prevailed between September 2011 and January 2015. Their work can be seen as a special case of our general procedure. This is for three reasons. First, our method can be applied to two-sided target zones, and not just two single-sided ones. Second, it is not restricted to the assumption that the underlying follows a geometric Brownian motion (GBM). Third, it applies to general exchange rate pairs and does not rely on the assumption that the domestic interest rate is consistently below the foreign one. Hanke et al. [79] had justifications for at least two of their three assumptions. The euro/Swiss-franc target zone that they studied is single-sided. This simplifies matters drastically by reducing (4.4) to (4.1) free of any mirror effects. Throughout the duration of that target zone policy, the euro interest rate was strictly above that of the Swiss franc, which was approximately zero. Since in that case it is never worth exercising the American put early, its price coincides with the European put. For convenience, the authors then made the additional assumption that the underlying fundamental follows a geometric Brownian motion. This allowed them to express the likelihood (4.18) in analytical form.

In our method, we do not rely on any of the above assumption. The fundamental needs not follow a GBM, the target zone can be single- or double-sided and the domestic interest rate may fluctuate arbitrarily above or below the foreign one. The transition probability  $p_F$  can only

---

<sup>12</sup> The Kolmogorov (forward) partial differential equation is also known as Fokker-Planck equation.

be expressed analytically for very special cases, e.g. for GBM. However, as we will discuss in more detail below, it can in general be approximated with arbitrary precision by expanding it into Hermite polynomials [85].<sup>13</sup> We use this expansion to approximate arbitrary functional shapes of  $\sigma(F; \theta)$ . The relationship between  $S$  and  $F$  and its inverse is calculated numerically as discussed in section 4.2. The derivative in (4.17) is approximated using the Richardson scheme [87]. This is all we need to estimate the parameters  $\theta$ , the implied maturity  $\tau$  and the fundamental  $F$  from the exchange rate data. In the next section, we apply this method to the USD/HKD target zone.

## 4.5 Application to the USD/HKD exchange rate

### 4.5.1 General set-up

Hong Kong experienced persistent inflows of funds during 2003 and 2005 driven by weakness of the US dollar, market speculations about the revaluation of the renminbi, and strong economic recovery in Hong Kong. In the absence of an explicit commitment by the Hong Kong Monetary Authority (HKMA) to sell Hong Kong dollars (HKD) on the strong side, the HKD appreciated abruptly from around 7.80 HKD to 7.70 HKD per US dollar (USD) in late 2003. To stabilize the exchange rate, the HKMA stepped in to conduct strong-side monetary operations. On May 18th, 2005, the [88] announced that it would conduct market operations to keep the exchange rate between 7.75 and 7.85 HKD per USD. This target zone has been active over the last decade as can be seen from the top plot of Figure 4.3.

Using the method presented in the previous section, we will now estimate the fundamental exchange rate value that would have prevailed in absence of the target zone. Specifically, we consider intraday exchange rate data between January 2010 and December 2016. We have resampled the data to 5 minute intervals ( $\Delta t_i = 5\text{min}$ ) and grouped the data into weeks. There are approximately 360 weeks from 2010 to 2016. Since the forex exchange market is open 24 hours a day, we have around  $24 \cdot 12 \cdot 5 = 1400$  exchange rate fluctuation datapoints  $\{\Delta S_i\}$  per week. For each week, we then estimate the volatility, implied credibility and fundamental value by maximizing the logarithm of (4.18).

To start, a parametric shape for the fundamental has to be specified, which we take as the

---

<sup>13</sup> There are some mild requirements on the smoothness of the process [86]. The volatility functions that we consider do not break these assumptions.

reasonably general CEV process<sup>14</sup>

$$dF = (r - \delta) F dt + \alpha F^\beta dW_t, \quad \alpha, \beta > 0. \quad (4.19)$$

As above, the parameters  $r$  and  $\delta$  denote the domestic and foreign risk free rate, respectively. We have used the daily 3-month Libor as a proxy. Both US and HK interest rates have been consistently below 1% (Figure 4.3). The parametrization  $\sigma = \alpha F^\beta$  is chosen because it encompasses important special cases, such as geometric Brownian motion ( $\beta = 1$ ) or the square-root process ( $\beta = 1/2$ ). Our method is not restricted to this choice, and any other parametrization is possible.

#### 4.5.2 Approximate transition probabilities

As a last crucial ingredient to the likelihood function (4.18), we require an approximate analytical expression for the transition density of stochastic process (4.19). More generally, consider a quantity  $F$  that evolves according to the stochastic process

$$dF = \mu(F) dt + \sigma(F) dW_t \quad (4.20)$$

for some given drift and volatility function  $\mu$  and  $\sigma$ . Denote by  $p_F(F | F_0, \Delta t)$  the probability density of being at position  $F$ , conditional on being at position  $F_0$  at time  $t = 0$ . Equivalently, we may also work with the density  $p_F(\Delta F | F_0, \Delta t, \theta)$  of the fundamental moving a step  $\Delta F$  within a timespan  $\Delta t$ , conditional on its current value  $F_0$ . This is of course just a shifted version of  $p_F$ . This transition density is obtained by solving the Kolmogorov-equation

$$\frac{\partial p_F}{\partial t} = -\frac{\partial}{\partial F} (\mu p_F) + \frac{1}{2} \frac{\partial^2}{\partial F^2} (\sigma^2 p_F). \quad (4.21)$$

While the partial differential equation (4.21) can be solved analytically only for some specific cases, it can be approximated reasonably well by expanding it into Hermite polynomials [85, 86], as we shall now briefly explain.

One starts by transforming the process for  $F$  into a process  $y$  with unit diffusion

$$dy = \mu_y(y) dt + dW_t \quad (4.22)$$

---

<sup>14</sup> The Constant Elasticity of Variance (CEV) model has been studied by [89].

through

$$\mu_y = \frac{\mu}{\sigma} - \frac{1}{2} \frac{\partial \sigma}{\partial F} \quad \text{and} \quad y = \int^F du / \sigma(u) \quad (4.23)$$

where any primitive of the function  $1/\sigma$  may be selected. Since the process has a standardized diffusion term, its transition density  $p_y(\Delta y | y_0)$  can be expanded into Hermite polynomials up to order  $K$  according to

$$p_y^{(K)}(\Delta y | y_0) = \frac{1}{\sqrt{\Delta y}} \phi\left(\frac{y - y_0}{\sqrt{\Delta y}}\right) \exp\left(\int_{y_0}^y dw \mu_y(w)\right) \sum_{k=0}^K c_k(y | y_0) \frac{\Delta^k}{k!} \quad (4.24)$$

with  $\phi(z) = \exp(-z^2/2) / \sqrt{2\pi}$  the normal Gaussian density,

$$c_j(y | y_0) = \begin{cases} 1 & \text{if } j = 0 \\ \frac{j}{(y - y_0)^j} \int_{y_0}^y dw (w - y_0)^{j-1} \left( \lambda_y(w) c_{j-1}(w | y_0) + \frac{1}{2} \frac{\partial^2 c_{j-1}}{\partial w^2} \right) & \text{if } j \geq 1 \end{cases} \quad (4.25)$$

and

$$\lambda_y(y) = -\frac{1}{2} \left( \mu_y^2 + \frac{\partial \mu_y}{\partial y} \right). \quad (4.26)$$

The approximation for  $p_F$  is finally obtained through the usual transformation rule for probability densities, similar to the one used in (4.17). For transformation (4.23), it is specifically given by  $p_F^{(k)} = p_y^{(k)} / \sigma$ . For the CEV process (4.19), the second order approximation then reads

$$p_F(y) \approx -\frac{1}{2} \log(2\pi\Delta t) - \log(\alpha F^\beta) + \frac{A}{\Delta t} + B + C \cdot \Delta t \quad (4.27)$$

with

$$y = \frac{F^{1-\beta}}{\alpha(\beta-1)}, \quad (4.28a)$$

$$y_0 = y(F_0) \quad (4.28b)$$

$$\mu = r - \delta \quad (4.28c)$$

$$A = -\frac{1}{2} (y - y_0)^2 \quad (4.28d)$$

$$B = \frac{1}{2} \frac{\beta \log(y/y_0)}{\beta - 1} - \frac{1}{2} (\mu (\beta - 1) (y - y_0) (y + y_0)) \quad (4.28e)$$

$$C = \frac{1}{6(y - y_0)} \left[ 3 \frac{(\beta - 2) \beta (y - y_0)}{4 (\beta - 1)^2 y y_0} - \frac{3 \mu (1 - \beta) (4 - 11\beta + 6\beta^2) (y - y_0)}{(\beta - 1) (3\beta - 4)} - \mu^2 (\beta - 1)^2 (y^3 - y_0^3) \right]. \quad (4.28f)$$

We have checked that the third order approximation does not add significant improvements over the second order approximation.

### 4.5.3 Technical considerations of the fitting procedure

The estimation of the parameters  $\alpha, \beta$  and  $\tau$  is challenging. The maturity dimension turns out to be sloppy [90,91] with respect to the dependence of the likelihood on  $\alpha$  and  $\beta$ . The dependence on  $\alpha$  and  $\beta$ , on the other hand, is strongly interdependent. This is easily understood: An increase in volatility can either be attributed to a larger  $\alpha$  or a larger  $\beta$ . The sloppiness of  $\tau$  together with the interdependence of  $\alpha$  and  $\beta$  results in a likelihood landscape that is difficult to handle by standard optimization schemes. We have therefore implemented a brute force optimization scheme that works with multiple layers. A first grid is put on the parameter space and the minimum is found by evaluating each grid point. The neighborhood around that minimum is then spanned with a finer grid, on which another minimum is found, and so on. The parameters on each of the approximately 360 trading weeks are then estimated in parallel on a supercomputer.

The computational effort for a single likelihood evaluation is considerable. We approximate the relation between  $S$  and  $F$  through  $S_n$  given by (4.5) with  $n = 100$ . The calculation of  $S$  alone requires therefore an evaluation of 200 American options. The options are priced with a binomial tree model as explained in section 4.2. Additionally, the relation between  $S$  and  $F$  has to be inverted and its derivative has to be taken. The evaluation of the transition probability  $p_F$  in terms of Hermite polynomials is fast since it can be pre-computed analytically as shown in (4.27). To optimize for performance, we have implemented the log-likelihood function in C++ and wrapped a Python function around it. This results in a duration of approximately 2 seconds per likelihood call.

To make sure that our estimates are reliable, we first test the method in a controlled environment. Using a Milstein scheme [92], we first generate an artificial fundamental value by simulating (4.19) in steps of 5 minutes for the business days between January 2010 and December 2016. We then calculate a synthetic exchange rate time series through  $S_{100}(F; \alpha, \beta, \tau)$  with  $S_{100}$  given by (4.5). Each week, we change the value of  $\alpha, \beta$  and  $\tau$  to make sure that the estimation works for different parameter configurations. For foreign and domestic risk free interest rate  $\delta$  and  $r$ , we choose the weekly averages of the 3 month Libor. The synthetic exchange rate is then passed to the brute force optimization function and  $\hat{\alpha}, \hat{\beta}, \hat{\tau}$  are estimated for each week by maximizing the logarithm of the likelihood (4.18). The estimated fundamental value is determined

by inverting  $S_{100}(F; \hat{\alpha}, \hat{\beta}, \hat{\tau})$ . The results were unsatisfactory when keeping both  $\alpha$  and  $\beta$  flexible. Since pricing American options with flexible  $\beta$  yields only marginal improvement over the simpler GBM case ( $\beta = 1$ ) [93], fixing  $\beta$  to 1 (or any other value) finally yields reliable estimates that are in good agreement with the true parameters, as is clearly visible in Figure 4.2. Despite the usage of a supercomputer and highly optimized code, we could not make the optimization grid fine enough to recover the true simulation parameters in under 24 hours. An alternative for future research is to speed up the evaluation of the likelihood by creating a lookup table for  $S_{100}$  over a specific range of parameters and then interpolate in-between. With these result we can move on to the analysis of the real data.

#### 4.5.4 Discussion of results

For each week between January 2010 and December 2016, we have estimated the volatility  $\alpha$  and implied maturity  $\tau$  for the USD/HKD exchange rate. We use the weekly average of the 3 month Libor to approximate foreign and domestic risk free rates  $d$  and  $r$ . The outcome is depicted in Figure 4.3. As can be seen from the inset, both risk free rates are consistently below 1%, but alternate in height which makes an approximation through European options invalid.

The estimated fundamental is almost exclusively below the target zone, indicating the strong effect the target zone has on the observable exchange rate. It also implies a significant risk to the economy in case the defense policy would be suddenly abandoned. Indeed, to avoid being arbitrated, the central bank can only abandon such a policy abruptly, as illustrated by the decision of the Swiss National Bank on January 15, 2015 that surprised everyone. Hong Kong businesses relying on exports would thus be advised to hedge against a sudden increase of the value of HKD for instance through future contracts.

Another interesting feature is the sequence of peaks in the implied maturity between 2013 and 2015, in coordination with times of sustained pressure on the Hong Kong dollar. This can be interpreted as follows. On the one hand, the continuous pressure signals the necessity for the central bank to continue imposing the target zone in order to keep the HKD from appreciating too much. This expectation is then interpolated into the future, enforcing the belief that the target zone policy will be held for a while. On the other hand, actively defending (the strong side of) a target zone can be costly for the central bank, as it requires purchasing huge amounts of a weakening foreign currency. While a sustained pressure always carries the danger of sudden abandonment, it may also prove the commitment of the central bank to keep its policy upright,



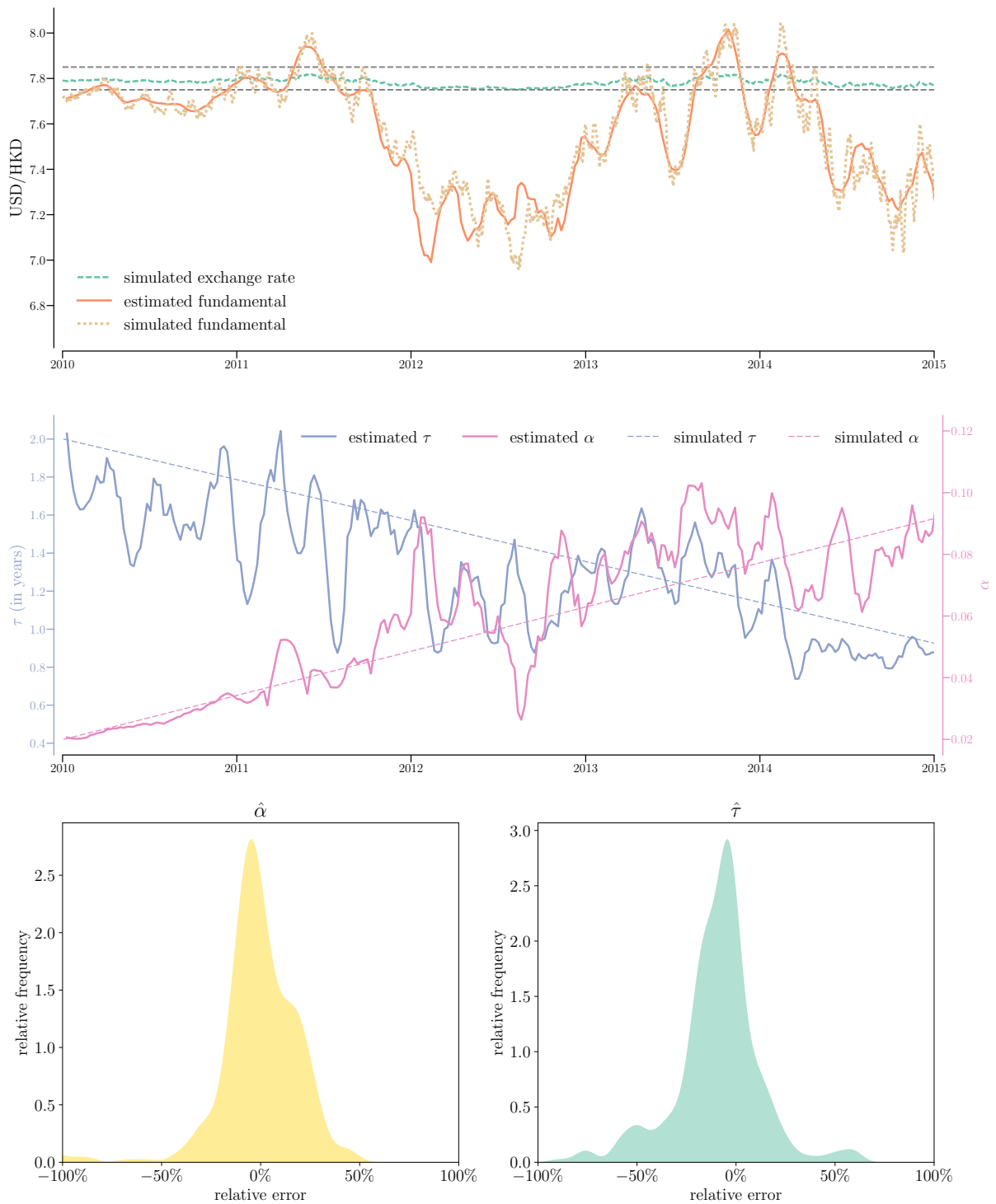


Figure 4.2: Result of estimating volatility  $\alpha$  and maturity  $\tau$  under controlled conditions. An artificial fundamental is generated by simulating (4.19) with weekly changing volatilities  $\alpha$  and fixed  $\beta = 1$ . The exchange rate is then determined as  $S_{100}(F; \alpha, \beta, \tau)$  with weekly changing implied maturities  $\tau$ . Given this exchange rate data, we estimate those parameters for each week by maximizing the logarithm of the likelihood (4.18). The fundamental is finally estimated by inverting  $S_{100}(F; \hat{\alpha}, \beta, \hat{\tau})$ . Although there are some fluctuations in the parameter estimates, it is readily visible that the underlying fundamental is reconstructed fairly well. To reduce smaller fluctuations, we have smoothened the estimated curves with a Savitzky-Golay filter. This does not alter the observation qualitatively but helps visualizing the trends.

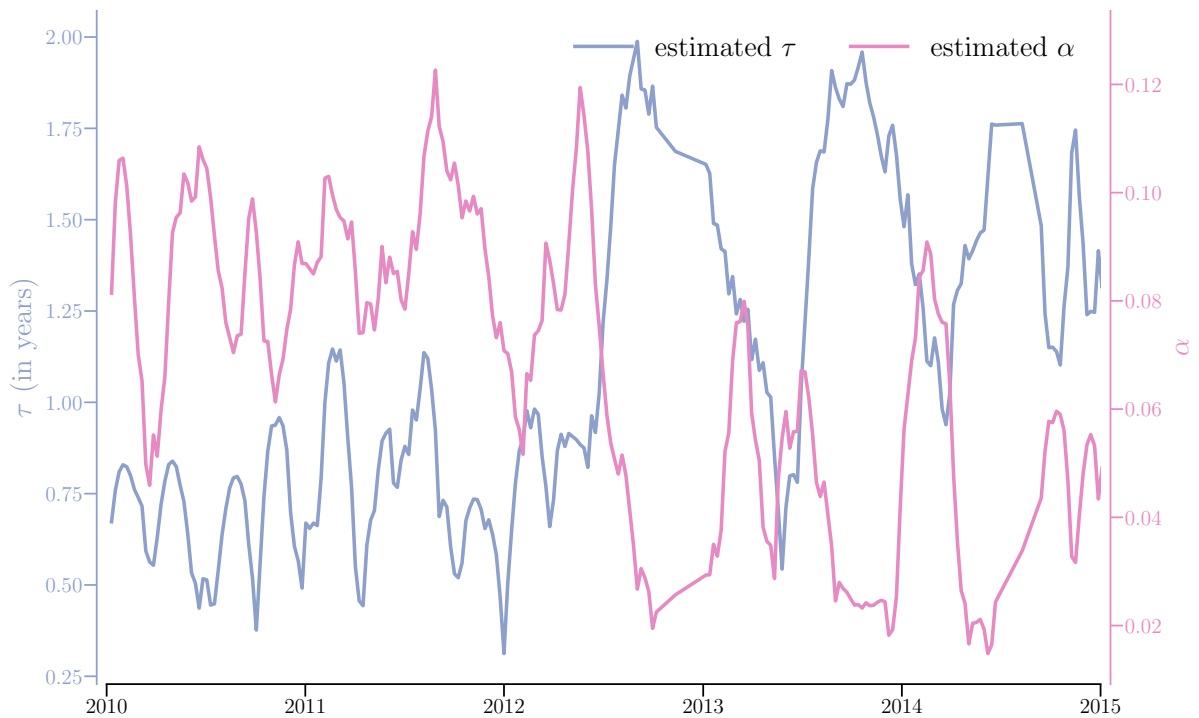
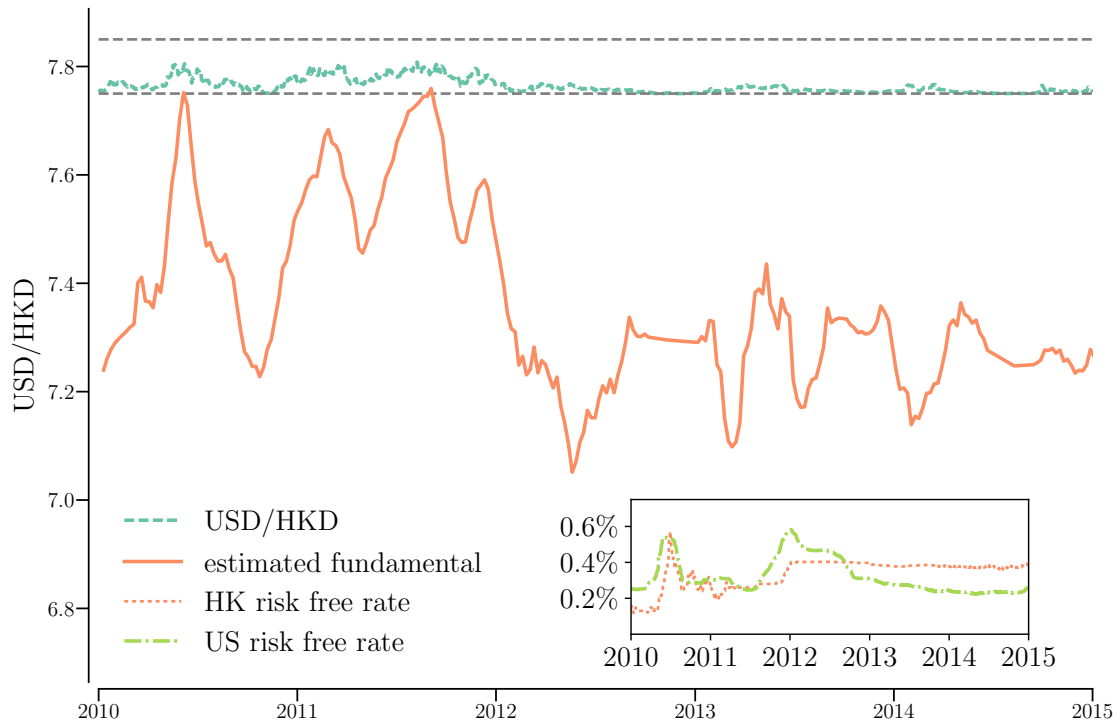


Figure 4.3: Result of fitting the model (4.19) to real USD/HKD exchange rate data with  $\beta = 1$ . The estimated volatility  $\hat{\alpha}$  and implied credibility  $\hat{\tau}$  have been obtained from maximization of the logarithm of the likelihood (4.18). The fundamental is estimated by inverting  $S_{100}(\hat{\alpha}, \beta, \hat{\tau})$ . To reduce fluctuations, we have smoothed the estimated curves with a Savitzky-Golay filter. This does not alter the observation qualitatively but helps visualizing the trends. The inset of the top Figures shows risk free rates approximated through weekly averages of the 3 month Libor.

thereby increasing the implied maturity. Our results indicate that here the later is the case. This is partially in line with results for the euro/Swiss franc target zone studied by Studer-Suter and Janssen [94]. They find that, while not perceived as credible at first, periods of sustained pressure and a continuous defense may actually strengthen the target zone credibility later on.

## 4.6 Conclusions

In this chapter we have presented a complete description of the dynamics of an exchange rate in the presence of a target zone in terms of iterated option prices. We have first made the idea intuitive in terms of single sided target zones. A target zone on the strong side is equal to holding the fundamental value and a long position in an American put option with strike price at that boundary. Similarly, a target zone on the weak side is equal to holding the fundamental and being short an American call option. This has lead naturally to the naive representation of the target zone in terms of one put and one call option. But just like placing two mirrors in front of one another produces an infinite number of reflections, we have shown subsequently that an infinite number of correction terms has to be taken into account to fully describe the target zone. Higher order correction terms are especially important in narrow target zones such as the USD/HKD exchange rate that has been studied in this chapter.

We have explained in detail how the nested option prices are calculated numerically, and the manifold technical challenges have been discussed. Albeit computationally involved, the mapping of target zones to option prices provides a convenient tool to analyze different target zone environments. We have argued how soft target zone boundaries, forward looking target zones and market inefficiencies can be expressed in the picture of option prices.

To give a flavor of such applications, we have subsequently applied maximum likelihood methods to estimated the implied option maturity. This maturity has a natural interpretation in terms of target zone credibility. Our method can be applied to general target zone environments and does not rely on specific assumptions about interest rate differentials or parametric shape of the fundamental price evolution.

The method has been applied to the USD/HKD target zone where the fundamental was obtained to be consistently below the lower boundary target zone, diagnosing a significant risk that the economy faces in case the target zone policy would be suddenly abandoned.



## Chapter 5

# An explicit mapping of currency target zone models to option prices <sup>†</sup>

In chapter 3 we have described forex target zones from a phenomenological perspective, whereas in chapter 4 we have relied on a close analogy to option pricing. After a short recapitulation of the two models, in this chapter, we will establish a quantitative link between these two approaches.

Currency target zones confine a country's domestic currency within a pre-specified upper and/or lower boundary with respect to some other currency. To achieve this, the central bank pledges to intervene and either buy or sell its currency, as soon as it touches the upper or lower boundary, respectively. Such target zones (TZs) have been employed frequently in the 1990s in preparation of the euro, and are still in use today. The TZ can be either single-sided, such as was the recent case for the Swiss franc - euro exchange rate (chapter 3), or two-sided, such as the Hong-Kong dollar - US dollar exchange rate (chapter 4).

Two prominent classes of modeling approaches have emerged from the literature. <sup>1</sup> One stream of models starts from a phenomenological equation in which the market's expectation of the central bank's policy is explicitly coupled back onto the price itself. The coupling parameter  $\gamma$  is a phenomenological constant that has to be determined from the data. Alternatively, TZs can be described as a portfolio made of a free float and of an additional long put (resp. short call) option position for the lower (resp. upper) TZ boundary. No phenomenological parameters are required here. Both approaches have their advantages. For instance, the phenomenological approach allows for a straightforward treatment of two-sided TZs. From an option perspective,

---

<sup>†</sup>This chapter is based on reference [12].

<sup>1</sup> We do not claim this classification to be exhaustive, and some other types of models are known. See also ref. [75] for an in-depth overview.

as discussed in the previous chapter, the treatment of two-sided TZs is much more involved, but it allows for flexible adjustment to different market conditions.

Until now, these two classes of models have been considered qualitatively equivalent, but no explicit quantitative mapping has been provided. In this chapter, restricting our attention to single-sided TZs for simplicity, we show that, in their basic form, there is no mathematical construction allowing one to map exactly one onto the other one. We provide the most parsimonious generalization of the phenomenological model, which allows for an exact mapping to the option approach, and which recovers the phenomenological model in the limit of small coupling parameter  $\gamma$ . A description of the basic phenomenological model is found in section 5.1, followed by the alternative option approach in section 5.2. A connection is derived in section 5.3 and its implications are discussed in section 5.4.

## 5.1 Phenomenological target zone models

Denote by  $s$  the logarithm of an exchange rate, i.e. the number of domestic currency units required to buy one unit of foreign currency. In a target zone arrangement, the central bank ensures that the exchange rate does not fall below some pre-defined value  $\underline{s}$  (a maximum  $\bar{s}$  is analogous). The phenomenological TZ model was pioneered by Krugman [40], who writes the intuitive equation

$$s = f + \gamma \frac{\mathbb{E}_t[ds]}{dt}, \quad (5.1)$$

with  $f$  the logarithm of the fundamental value that would prevail in a free market. The expected change of  $s$  is represented by the term  $\mathbb{E}[ds]/dt$ , where the phenomenological  $\gamma$  controls the sensitivity of  $s$  to the market's expected change of  $s$ . In absence of a target zone, it holds that  $\mathbb{E}[ds] = 0$  and hence  $s = f$ , meaning that the exchange rate reflects the value of the underlying, as determined by the efficient market. For  $s \gg \underline{s}$  one can approximate  $s \approx f$ . But as  $s$  approaches the boundary, the central bank is expected to intervene and defend the exchange rate band. Since the target zone is assumed credible, market participants act on this anticipation and put  $s$  back above the lower boundary themselves, without the central bank having to intervene actively ('honeymoon effect'). In case such endogenous market self-corrections are insufficient, the central bank steps in as a last resort by increasing money supply. Assuming that  $f$  follows a random

walk with volatility  $\sigma$ , equation (5.1) is solved with basic stochastic calculus, giving

$$s = f + \frac{1}{\rho} e^{\rho(\underline{f}-f)} \quad (5.2)$$

where  $\rho = \sqrt{2/(\gamma\sigma^2)}$  and  $\underline{f}$  is defined through  $s(\underline{f}) = \underline{s}$ , i.e.  $\underline{f} = \underline{s} - 1/\rho$ . The result is depicted in Figure 5.1. See chapter 3 for details. Various extensions of the model exist, such as imperfect credibility, soft target zone boundaries, forward looking TZs, and so on. See references in the introduction section and previous chapters for more details.

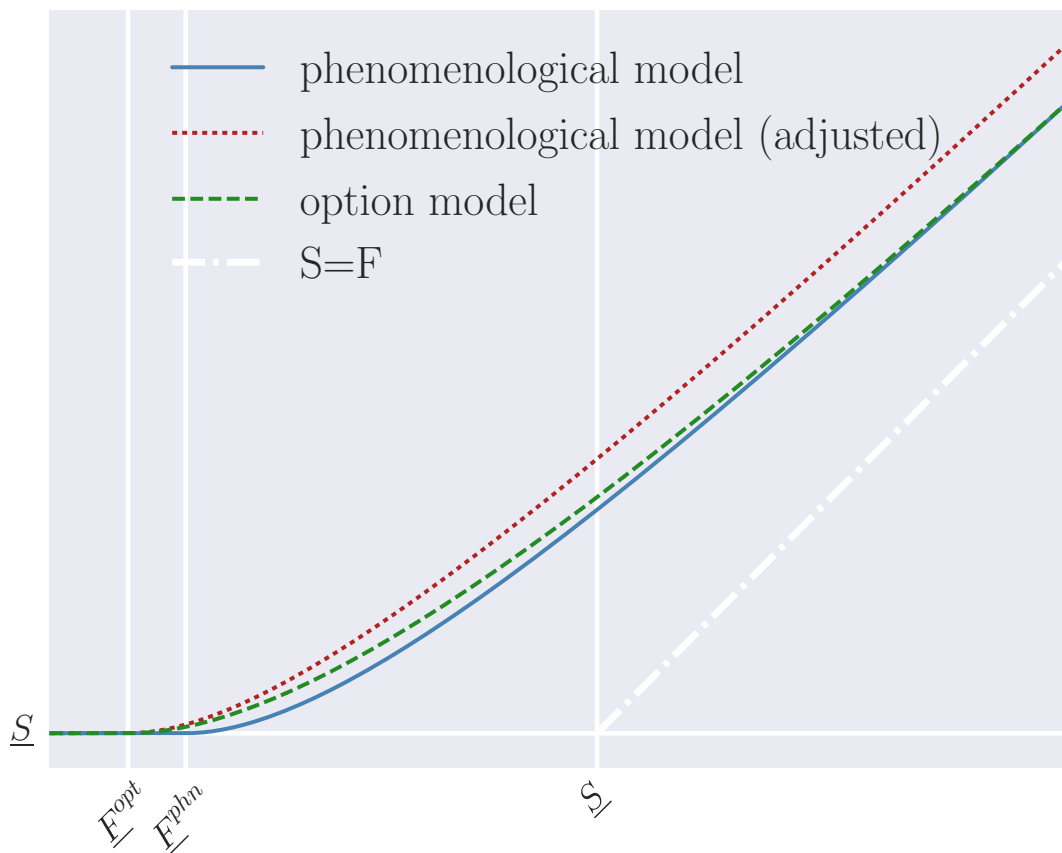


Figure 5.1: Comparison of the two target zone representations. The continuous line shows the (exponential of) the phenomenological approach (5.2), with  $\underline{f}^{\text{phn}} = \underline{s} - 1/\rho$  and  $\underline{F}^{\text{phn}} = \exp(f^{\text{phn}})$ . The dashed line is the alternative representation in terms of put options (5.3), with  $\underline{F}^{\text{opt}}$  given by  $\underline{S}\theta/(\theta - 1)$ . Here, we have used  $\underline{S} = 1.2$ , as well as annual  $\sigma = 20\%$ ,  $r = \delta = 1\%$  and  $\gamma$  is fixed through (5.24) to  $\approx 30$ . Generalizing the phenomenological model (5.1) through equations (5.19) and (5.22), the two models are found to coincide. The dotted line shows the phenomenological model, but with  $\rho$  adjusted such that  $\underline{F}^{\text{phn}}$  coincides with  $\underline{F}^{\text{opt}}$ . We can see that this naive adjustment does not make the two models coincide. The dash-dotted line is the linear relationship that prevails in absence of a target zone. We can clearly see that  $S \geq F$ , indicating the ‘repelling’ self-regulating effect of the credible target zone.

## 5.2 Target zones from an option perspective

As recalled above, the phenomenological models are traditionally formulated in logarithms of the involved quantities. In the context of options, it is more intuitive to work with the quantities themselves. We thus define  $S \equiv e^s$  and  $F \equiv e^f$ . If there is no TZ, one unit of foreign currency is worth  $S = F$  units of domestic currency, where the fundamental value  $F$  is determined by the free market. A central bank announcing that, no matter what the value of  $F$ , the exchange rate will always be held above  $\underline{S}$  is equivalent to holding one unit of the free floating currency and an additional American put option  $\mathcal{P}$  with underlying  $F$  and strike price  $\underline{S}$ ,<sup>2</sup>

$$S = F + \mathcal{P}(F, \underline{S}). \quad (5.3)$$

The option is American because it can be exercised at any time. Within the basic phenomenological model (5.1), the TZ is assumed to be installed ‘forever’. In the jargon of equation (5.3), this translates into the put having infinite maturity time.<sup>3</sup> Such options are called perpetual, and they can be priced analytically [95],

$$\mathcal{P}(F, \underline{S}) = \begin{cases} \underline{S} - F & \text{if } F \leq \underline{F} \\ \alpha F^\theta & \text{if } F > \underline{F} \end{cases} \quad (5.4)$$

with

$$\alpha = \frac{\underline{S}}{1 - \theta} \left( \frac{\theta \underline{S}}{\theta - 1} \right)^{-\theta} \quad (5.5)$$

and

$$\theta = \frac{\delta - r + \frac{\sigma^2}{2} - \sqrt{2r\sigma^2 + \left(\delta - r + \frac{\sigma^2}{2}\right)^2}}{\sigma^2}, \quad (5.6)$$

where  $r$  and  $\delta$  denote domestic and foreign risk free rate, respectively. Note that  $\theta$  is negative and  $r, \delta > 0$  is assumed. The value  $\underline{F}$  at which  $S$  touches the barrier is given by  $\underline{S}\theta/(\theta - 1)$ , and  $\sigma$  is again the volatility of the (log-) underlying. See Figure 5.1 for a visualization.

<sup>2</sup> Similarly, an upper boundary is as if one is short a call option.

<sup>3</sup> Within the option pricing framework, there is no need to assume infinite maturity. Instead, the ‘implied maturity’ can be fitted from the data and is interpreted as the market perceived target zone credibility. See for instance [13, 79]. Extensions to account for bounded credibility have also been developed in the phenomenological literature, e.g. by [44].



### 5.3 Relating the two approaches

Comparing the two solutions, the phenomenological solution (5.2) and the option version (5.3), we can see both from their mathematical structure as well as from Figure 5.1 that the two solutions are qualitatively similar but not equivalent. In order to make the two approaches consistent, we propose a generalized phenomenological equation of the form

$$s = f + g \left( \frac{\mathbb{E}_t[ds]}{dt} \right), \quad (5.7)$$

where  $g$  is some function that we are now going to determine. As a guide, we would like that  $g(u) \propto u$  for small  $u$ , so that Krugman's model could be seen as a first-order expansion of a more general theory. We are not going to impose this condition in our derivation, but will use it as an ex-post sanity check.

Applying Itô's lemma to  $ds = ds(f)$  in (5.7) and using the fact that a Wiener process has zero expectation value, we arrive at

$$s = f + g \left( \sigma \frac{\partial^2 s}{\partial f^2} \right). \quad (5.8)$$

Replacing the logarithmic quantities  $s$  and  $f$  by  $S$  and  $F$ , expression (5.8) yields

$$\log \left( \frac{S}{F} \right) = g \left( \frac{\sigma^2}{2} \frac{\partial^2 \log S}{\partial \log F^2} \right). \quad (5.9)$$

Applying the chain and product rule to some arbitrary function  $\zeta(F)$  gives

$$\frac{\partial^2 \zeta}{\partial \log(F)^2} = \frac{\partial}{\partial \log(F)} \left( \frac{\partial \zeta}{\partial \log(F)} \right) \quad (5.10a)$$

$$= \frac{\partial}{\partial \log(F)} \left( F \frac{\partial \zeta}{\partial F} \right) \quad (5.10b)$$

$$= \frac{\partial F}{\partial \log(F)} \frac{\partial \zeta}{\partial F} + F \frac{\partial}{\partial \log(F)} \frac{\partial \zeta}{\partial F} \quad (5.10c)$$

$$= F \frac{\partial \zeta}{\partial F} + F^2 \frac{\partial^2 \zeta}{\partial F^2}. \quad (5.10d)$$

Identifying  $\zeta \equiv \log(S)$ , a direct consequence of this identity is

$$\frac{\partial^2 \log(S)}{\partial \log(F)^2} = F \frac{\partial \log(S)}{\partial F} + F^2 \frac{\partial^2 \log(S)}{\partial F^2} \quad (5.11a)$$

$$= \frac{F}{S} \frac{\partial S}{\partial F} + F^2 \frac{\partial}{\partial F} \left( \frac{\partial}{\partial F} \log(S) \right) \quad (5.11b)$$

$$= \frac{F}{S} \frac{\partial S}{\partial F} + F^2 \frac{\partial}{\partial F} \left( \frac{1}{S} \frac{\partial S}{\partial F} \right) \quad (5.11c)$$

$$= \frac{F}{S} \frac{\partial S}{\partial F} + \frac{F^2}{S} \frac{\partial^2 S}{\partial F^2} - \frac{F^2}{S^2} \frac{\partial S}{\partial F} \quad (5.11d)$$

$$= \left( \frac{F}{S} - \frac{F^2}{S^2} \right) \frac{\partial S}{\partial F} + \frac{F^2}{S} \frac{\partial^2 S}{\partial F^2}, \quad (5.11e)$$

such that

$$\log \left( \frac{S}{F} \right) = g \left( \frac{\sigma^2}{2} \left( \frac{F}{S} - \frac{F^2}{S^2} \right) \frac{\partial S}{\partial F} + \frac{\sigma^2}{2} \frac{F^2}{S} \frac{\partial^2 S}{\partial F^2} \right). \quad (5.12)$$

From section 5.2, we know that, in the option approach,  $S = F + \alpha F^\theta$ . Plugging this solution into (5.12) leads to

$$\log \left( 1 + \alpha F^{\theta-1} \right) = g \left( \frac{\sigma^2}{2} \theta^2 + \frac{\sigma^2}{2} \frac{(1 - \theta - \theta^2)}{1 + \alpha F^{\theta-1}} + \frac{\sigma^2}{2} \frac{(\theta - 1)}{(1 + \alpha F^{\theta-1})^2} \right). \quad (5.13)$$

We observe that the  $F$  dependence appears in all terms in the form  $1 + \alpha F^{\theta-1}$ . This motivates the introduction of the following new variables:

$$x \equiv 1 + \alpha F^{\theta-1} \quad A \equiv \frac{\sigma^2}{2} \theta^2 \quad (5.14a)$$

$$B \equiv \frac{\sigma^2}{2} (1 - \theta - \theta^2) \quad C \equiv \frac{\sigma^2}{2} (\theta - 1), \quad (5.14b)$$

giving

$$\log x = g \left[ A + \frac{B}{x} + \frac{C}{x^2} \right]. \quad (5.15)$$

Next, we write without loss of generality  $g(\cdot) \equiv \log h(\cdot)$ , thus defining the function  $h(\cdot) > 0$ .

This leads to the functional equation

$$x = h \left[ A + \frac{B}{x} + \frac{C}{x^2} \right]. \quad (5.16)$$

We define  $u \equiv A + \frac{B}{x} + \frac{C}{x^2}$  and solve for  $x$  as a function of  $u$ :

$$x = -\frac{B}{2(A-u)} \pm \frac{\sqrt{B^2 - 4AC + 4Cu}}{2(A-u)}. \quad (5.17)$$

As becomes immediately clear below, only the solution with ‘+’ is consistent. Thus,

$$h(u) = -\frac{B}{2(A-u)} + \frac{\sqrt{B^2 - 4AC + 4Cu}}{2(A-u)}. \quad (5.18)$$

Using (5.15), we have thus found that the generalization of the phenomenological equation (5.1) into (5.7) with

$$g(u) = \log \left( -\frac{B}{2(A-u)} + \frac{\sqrt{B^2 - 4AC + 4Cu}}{2(A-u)} \right) \quad (5.19)$$

leads to the same solution as the option approach.

While, technically, expression (5.19) presents the formal result of this chapter, it is insightful to further expand this expression into a Taylor series around  $u = 0$ . This is justified because the argument of the function  $g(\cdot)$  is  $\mathbb{E}[ds]/dt$ , which is continuously connected to 0, and the expected change at any instance is usually small. The Taylor expansion of  $h(u) = \exp(g(u))$  given by expression (5.18) leads to

$$h(u) = 1 + \frac{2}{\sigma^2} \frac{1}{1 + \theta(\theta - 1)} u + \mathcal{O}(u^2). \quad (5.20)$$

We can then define

$$\gamma \equiv \frac{2}{\sigma^2} \frac{1}{1 + \theta(\theta - 1)}, \quad (5.21)$$

and obtain <sup>4</sup>

$$s = f + \log \left( 1 + \gamma \frac{\mathbb{E}[ds]}{dt} + \mathcal{O} \left( \left[ \frac{\mathbb{E}[ds]}{dt} \right]^2 \right) \right). \quad (5.22)$$

Since  $\log(1+x) \sim x$  for  $x$  small, this expression recovers the original approach (5.1) to first-order. We interpret the logarithmic correction in the sense of a concave risk-averse ‘utility’ adjustment. Expression (5.22) thus provides a principled generalisation of the phenomenological Krugman model, anchored on the exact option approach.

## 5.4 Implications of the mapping

A practically important consequence of this mapping is that we can relate the phenomenological  $\gamma$  to economic quantities through (5.21). However, estimating  $\gamma$  from the data is very difficult as there are severe multicollinearity problems between the parameters  $\gamma$  and  $\sigma^2$ , that prevents

---

<sup>4</sup> If we had taken the other solution in (5.17), the zeroth order approximation in (5.20) is equal to  $(\theta-1)/\theta^2 \neq 1$ , which is inconsistent with the limit  $s \rightarrow f$  for  $\mathbb{E}[ds]/dt \rightarrow 0$ .

precise estimation [96]. Our mapping removes the need to estimate  $\gamma$ , as it is now a well-defined and explicit function of  $r, \delta$  and  $\sigma$ . It must be noted that  $\sigma$  itself is unobservable, since  $f$  is unobservable (only  $s$  is the observable exchange rate). But compared to the rather ad hoc postulated  $\gamma$ , it is much more tangible as a value, and can be estimated more accurately through fundamental analysis or likelihood methods developed in the previous chapter.

To gain more insights into our result (5.21), we consider the special case  $r = \delta$  (domestic and foreign interest rate are equal). In that case, the  $\theta$  exponent (5.6) simplifies into

$$\theta = \frac{1}{2} - \frac{1}{2} \sqrt{1 + \frac{8r}{\sigma^2}}, \quad (5.23)$$

and we obtain

$$\gamma^{-1} = r + \frac{\sigma^2}{2}. \quad (5.24)$$

In the phenomenological solution (5.2),  $1/\rho = \sigma\sqrt{\gamma/2}$  is the typical amplitude of the variation of the logarithm of the fundamental value over which the non-linear relation between  $s$  and  $f$  can be observed. If we assume, for instance, that annual volatility  $\sigma$  and domestic risk free rate  $r$  are of the order of 20% and 3%, respectively, we find  $1/\rho \approx 3\sigma$ . In other words, if the TZ is perfectly credible, the non-linear relationship is observable over as much as three annual volatilities away from the target zone boundary.

An examination of expression (5.24) rationalises why estimates of  $\gamma$  vary so much in the literature, ranging from  $\gamma = 0.01$  to 100 [70, 96, 97]. Indeed,  $\gamma$  is approximately the inverse of small numbers  $r$  and  $\sigma^2$  and small errors in their estimations amplify in large uncertainties. Moreover, in the regime of small volatility of the fundamental factor compared with interest rate differentials such that  $\sigma^2 < r - \delta$ , the exponent  $\theta$  is typically large negative ( $\theta \simeq \frac{\delta-r}{\sigma^2}$  to first-order), and thus  $\gamma \simeq 2 \left( \frac{\sigma}{r-\delta} \right)^2$  can be very small. Hence,  $\gamma$  varies by many orders of magnitudes when going from  $\sigma^2 < r - \delta$  to  $\sigma^2 > r - \delta$ .

In addition, it is unrealistic to assume that the put option in (5.3) is perpetual. In practice, the target zone is not infinitely credible, and hence, the representative put option should have a maturity. That ‘implied’ maturity, which can be interpreted as the market’s perceived duration of the current TZ policy, can be determined from the observable exchange rate. With a maturity parameter, the rather wide power-law shaped ATM (at-the-money) regime of the perpetual put (5.4) is significantly shortened, as the price of the put is replaced by the well-known Black-Scholes formula of the American put. With the help of additional assumptions on the credibility

of the TZ, this opens up the possibility of quantifying further the width of the ATM regime, and therefore of estimating  $\gamma$  via a similar mapping as done above, replacing the perpetual option pricing formula (5.4) by the corresponding relevant option price. We leave this for future work.

## 5.5 Conclusions

In this chapter, we have shown how the phenomenological TZ model of Krugman [40] discussed in chapter 3 has to be generalized in order to recover the alternative representation through option prices from chapter 4. This opens the possibility of resolving the difficult estimation of  $\gamma$  through that of economic fundamentals from which it derives. This mapping further highlights that Krugman's assumption of perfect credibility is unrealistic in the sense that it over-estimates the width over which the non-linear relationship between  $s$  and  $f$  prevails.



## Chapter 6

# Evidence of a bimodal US GDP growth rate distribution: A wavelet approach <sup>†</sup>

In the previous chapters, we have discussed free random walks that are constrained by a rigid barrier. In the next section, we will see another form of a constrained random walk: GDP growth rates described as Lévy flights in steep potentials. One key ingredient to that theory will be the bimodal structure of (US) GDP growth rates. In this chapter, we set the scene by providing thorough statistical evidence that the growth rate distribution is indeed bimodal.

The dynamics of the growth of GDP (gross domestic product) is arguably the most scrutinised metric quantifying the overall economic development of an economy. A weak annual growth rate of GDP, as has been characterising the US and Europe in the years following the financial crisis of 2008, is interpreted as underperformance, which has called for unorthodox monetary policies attempting to fix it [98]. In contrast, a strong growth of GDP is usually lauded, because it reflects a rise of living standards and is generally accompanied by decreasing unemployment. But what is meant by ‘weak’ or ‘strong’ growth? Is there a ‘natural’ growth rate? Does past growth rates of GDP imply future growth rates? This last question is particularly relevant in the present context of small growth compared with previous decades in developed countries and the argument by many that we may have shifted to a ‘new normal’ of slower intrinsic growth [99].

It is well-known that plotting the logarithm of the real US GDP per capita over the last one hundred years looks remarkably linear with a slope estimated between 0.019 and 0.02. In other words, the inflation adjusted GDP per capita exhibits a long term average growth of 1.9 – 2% per year, see the dashed linear fitted lines in Figure 6.3. The occurrence of such a near

---

<sup>†</sup>This chapter is based on reference [13].

trend-stationary long run growth covering a period with two world wars, the cold war and its associated proxy wars, the collapse of the Bretton Woods System in 1973, several large bubbles, crashes and recessions and strong changes in interest rate policies, is truly remarkable. It entices one to entertain the possibility of an equilibrium or natural growth rate, which then could be extrapolated in the future. Business cycles would then be viewed as fluctuations around this equilibrium growth rate.

In this chapter, we challenge the standard hypothesis that business cycles are merely fluctuations (or transient deviations) around a stationary equilibrium growth rate. By analyzing quarterly data of real US GDP per capita (r-US-GDP-pc) between 1947 until 2015, we find both parametric and non-parametric evidence that the GDP growth rate density is bimodal, with peaks at a high and a low growth rate of  $\rho_{\text{high}} \approx 3\%$ , and  $\rho_{\text{low}} \approx 1\%$ , respectively.<sup>1</sup> This leads to the conclusion that the US economy per capita is intrinsically composed of alternation between regimes of strong growth rate  $\rho_{\text{high}}$ , associated with booms (or bubbles), and regimes of low growth rate  $\rho_{\text{low}}$  that include plateaus (and recessions). Alternations between those two regimes give rise to business cycles. Only when viewed at larger scales, these two alternating regimes renormalize to an effective long-term growth rate  $\rho_{\text{lt}} \approx 2\%$  that is between  $\rho_{\text{low}}$  and  $\rho_{\text{high}}$ . Throughout this chapter, we stick to quarterly data from 1947 until 2015. In the next chapter, we show a similar analysis for annual GDP data from 1800 until today.

Our findings have important economic and policy implications. The existence of a well-characterised strong growth regime with average growth rate  $\rho_{\text{high}}$  often leads to the misleading expectations that it is the normal that reflects a well-functioning economy, while the other mode of low growth  $\rho_{\text{low}}$  is considered abnormal, often interpreted as due to a surprising shock, bringing considerable dismay and pain, and leading to policy interventions. Our finding of a robust bimodal distribution of GDP growth rates over the whole history of the US suggests that this interpretation is incorrect. Rather than accepting the existence of the long-term growth rate as given and interpreting the deviations from it as perturbations, the bimodal view of the GDP growth suggests a completely different picture. In this representation, the long-term growth rate is the result of a subtle compensation between the high and low growth regimes that alternate continuously. The overall growth dynamics that emerges is that the US economy is growing in a punctuated way [100], following phases of strong growth that are intrinsically unsustainable, followed by corrections or consolidations until the next boom starts. In other

---

<sup>1</sup> Unless stated otherwise, we refer from hereon to the real GDP per capita (r-US-GDP-pc) simply as GDP.



words, the approximately long-term growth rate reflects an economy that oscillates between booms and consolidation regimes. Because of the remarkable recurrence of the strong regime and in view of its short-term beneficial effects, economists and policy makers are tempted (and actually incentivised) to form their expectations based on it, possibly catalysing or even creating it in a self-fulfilling prophecy fashion even when the real productivity gains are no more present, as occurred in the three decades before the 2008 crisis [101].

We suggest that the transient strong growth regimes can be rationalised within the framework of the ‘social bubble hypothesis’ [102–105], in the sense that they result from collective enthusiasm that are similar to those developing during financial bubbles, which foster collective attitude towards more risk taking. The social bubble hypothesis claims that strong social interactions between enthusiastic supporters weave a network of reinforcing feedbacks that lead to widespread endorsement and extraordinary commitment by those involved, beyond what would be rationalised by a standard cost-benefit analysis. For a time, the economy grows faster than its long-term trend, due to a number of factors that reinforce each other, leading to a phase of creative innovation (e.g. the internet dotcom bubble) or credit based expansion (e.g. the house boom and financialisation of the decade before 2008). These regimes then unavoidably metamorphose into a ‘hangover’, the recovery and strengthening episode until the next upsurge.

Despite the huge attention paid to the analysis of GDP growth rate fluctuations, this work presents, to the best of our knowledge, for the first time a detailed analysis of the bimodal nature of its density distribution. In the next section, we first address the question of bimodality from the perspective of classic, parametric business cycle models. This yields first evidence of a bimodal distribution density. However, the reported results are susceptible to small changes in parameters and noise. Subsequently, we thus turn to a non-parametric analysis. The usage of wavelets as adequate tool for this analysis is motivated in section 6.2. Section 6.3 presents the wavelet methodology itself. Results are reported in 6.4. Section 6.5 is focused on the evidence supporting the bimodal structure of the distribution of GDP growth rates and section 6.6 concludes. The next chapter complements the presentation with annual GDP data over the past 200 years.

## 6.1 Business cycles as Markov autoregressive processes

First introduced by Hamilton [106], Markov autoregressive processes have become a popular tool for GDP business cycle analysis. In standardized form, such a process is written as

$$\rho_t = \mu(S_t) + \sum_{i=1}^n \phi_i \cdot (\rho_{t-i} - \mu(S_{t-i})) + \epsilon_t \quad (6.1)$$

with  $\rho_t$  the growth rate at time  $t$ ,  $S_t$  describes the regime of the economy at time  $t$ , which can take  $m$  different states that are reflected in  $m$  different values  $\mu_1, \dots, \mu_m$  for  $\mu(S_t)$ . The parameters  $\phi_1, \dots, \phi_n$  specify the autoregressive characteristics of the process and  $\epsilon_t$  is Gaussian noise with zero mean and standard deviation  $\sigma$ . The switching from a regime  $i$  to a regime  $j$  at any time step is determined by a matrix of transition probabilities  $p_{ij}$ . In the classic Hamilton [106] model,  $m = 2$  and  $n = 4$  and we shall stick here to this choice. The main idea is that there are two states of the economy, a boom state  $S_t = 1$  and a recession state  $S_t = 0$ . From time  $t$  to  $t+1$ , the economy switches between these two states with probabilities  $p_{01}$  and  $p_{10}$ , respectively. There is a debate about the nature of the switching probabilities  $p_{ij}$ . Hamilton [106] assumes that the probabilities are independent of the business cycle duration whereas others assume that there is a dependence. A famous model of the second type is by Durland and McCurdy [107]. Formal tests [108] lead to the concludes that Hamilton's [106] assumption of duration independence is legitimate. We therefore stick to the model of Hamilton to keep the discussion simple.

What is the prediction of these models for the GDP density distribution? In order to answer this question, we first fit the 9 parameters of the model (6.1) to the 270 datapoints of quarterly r-US-GDP-pc from 1947 to 2015.<sup>2</sup> Using these parameters, we then simulate a synthetic time series over a period of 100,000 years and extract the asymptotic density distribution from the simulated growth rates.

When fitting (6.1) to the raw GDP data, using the same data set of initially used by Hamilton [106], we recover the same parameters, which are also close to those obtained for the expanded data from 1947 to 2015. However, the analysis of the GDP density distribution reveals a feature that has been underestimated, namely the fact that the model anchors on the few large negative outliers present in the data. Fitted parameters are reported in Figure 6.1. The maximum likelihood algorithm identifies the few very negative outliers as one regime ( $\mu_0 = -1.7$ , i.e.  $-1.7\%$  quarterly, or  $-6.8\%$  yearly growth rate), and the average  $2\%$  yearly growth rate as the

---

<sup>2</sup> The following parameters have to be fit:  $\mu_0 = \mu(S_t = 0)$ ,  $\mu_1 = \mu(S_t = 1)$ ,  $\phi_1, \phi_2, \phi_3, \phi_4, p_{12}, p_{21}$  and  $\sigma$ .

second regime ( $\mu_1 = 0.5$ , i.e. 0.5% quarterly growth). The recession regime is interpreted by the calibration as a rare event (transition probability from average growth regime to recession regime  $\approx 3\%$ ) that immediately jumps back to the normal average growth of roughly 2% yearly growth. This anchoring on the few large negative quarters is clearly revealed by the synthetic GDP density distribution shown in Figure 6.1c.

To obtain results that are more in line with actual business cycle fluctuations, we discard all datapoints below  $-4\%$  and above  $8\%$  yearly growth (below  $-1\%$  and above  $2\%$  growth per quarter). This specific choice for the cut-off range is ad hoc and determined through simple visual inspection. The results are robust to reasonable variations of these two thresholds. Moreover, we overcome this arbitrariness with non-parametric estimates in the subsequent sections. Without the outliers, the maximum likelihood estimates of model (6.1) converge to more balanced results. The boom and recession regimes are now identified as  $0.8\%$  and  $3.2\%$  yearly growth, respectively. As is clearly visible in Figure 6.1d, the corresponding stationary growth rate distribution is bimodal. This is in good agreement with non-parametric evidence reported in the subsequent sections. We also observe that the left (recession) peak is larger than the right (boom) peak, indicating an imbalance between the two.

However, the results of this section have to be interpreted with caution, as there may be multiple local optima of the likelihood function. An inspection of the likelihood function (not shown here) also reveals that there are sloppy manifolds in the 9-dimensional parameter space, making the calibration intrinsically difficult with several degenerate solutions (see ref. [109] and references therein). We thus turn our attention to an alternative, parameter free approach in the next section.

The results of this section are summarized in Figure 6.1. For comparison, we also show the density distribution corresponding to the original fits obtained by Hamilton [106] and by Durland and McCurdy [107]. Interestingly, the two models differ in their prediction: The Hamilton model converges to a unimodal distribution, whereas the model by Durland and McCurdy yields a more bimodal structure. It is interesting to see that the two models predict different shapes. One should be cautious in comparing the results of these two models with ours, since they have analyzed nominal GNP data, and not real GDP per capita. We use real GDP per capita since it represents a measure for real innovation and productivity gains. In contrast, the total nominal US GDP contains two additional contributions to its growth: population growth and inflation, both of which do not represent increased income per individual.

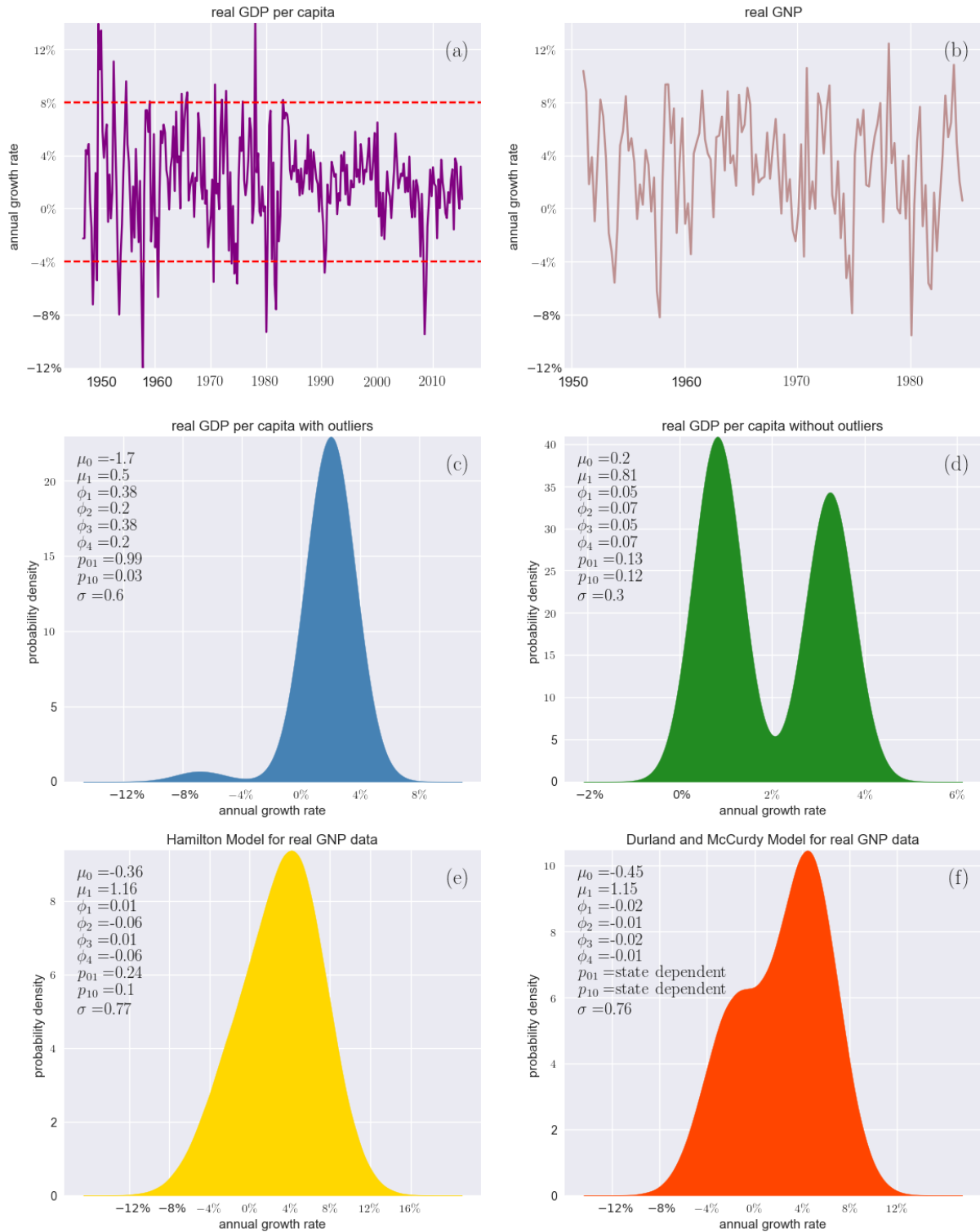


Figure 6.1: Summary of all results discussed in section 6.1. For easier interpretation, all quarterly growth rates in this figure have been rescaled to annual values, i.e. multiplied by a factor of four. The top two figures depict quarterly data of real GDP per capita and real GNP growth rates, respectively. The red dashed lines in figure (a) indicate the ad hoc definitions of outliers. For the creation of figure (d), only growth rates within that band have been fitted to the model (6.1). The resulting asymptotic growth rate distribution is clearly bimodal. This is in contrast to figure (c), where no outliers have been discarded. The bulk of all probability mass forms a unimodal structure, with a small additional hump around  $-8\%$  for the negative outlier regime. Figures (e) and (f) complement the analysis by showing the predictions of Hamilton [106] and Durland and McCurdy [107]. While the first distribution is clearly unimodal, the second one exhibits a bimodal structure.

## 6.2 Scale free business cycles

In order to examine further the modal structure of the USD growth rate distribution, we turn our attention to non-parametric models. The classification of economic growth into phases of booms and busts has been extensively investigated in the business cycle literature [110, 111].

Parametric or non-parametric, the traditional business cycle literature is concerned with pinning down explicit dates indicating a turning point in an economy. Furthermore, such approaches are often constrained by a priori imposed minimum or maximum business cycle durations, as specified for instance by the NBER business cycle committee [112]. Here, we avoid this approach of preconditioning and let the data decide what are the possible cycles that stand out and justify the identification of a cycle. A first parameter free approach to this problem is to perform a spectral analysis. Figure 6.2 shows the spectral density  $P(f)$  of the  $\ln(\text{r-US-GDP-pc})$  for both quarterly data since 1947 and annual data since 1800. We observe a scale-free continuum of

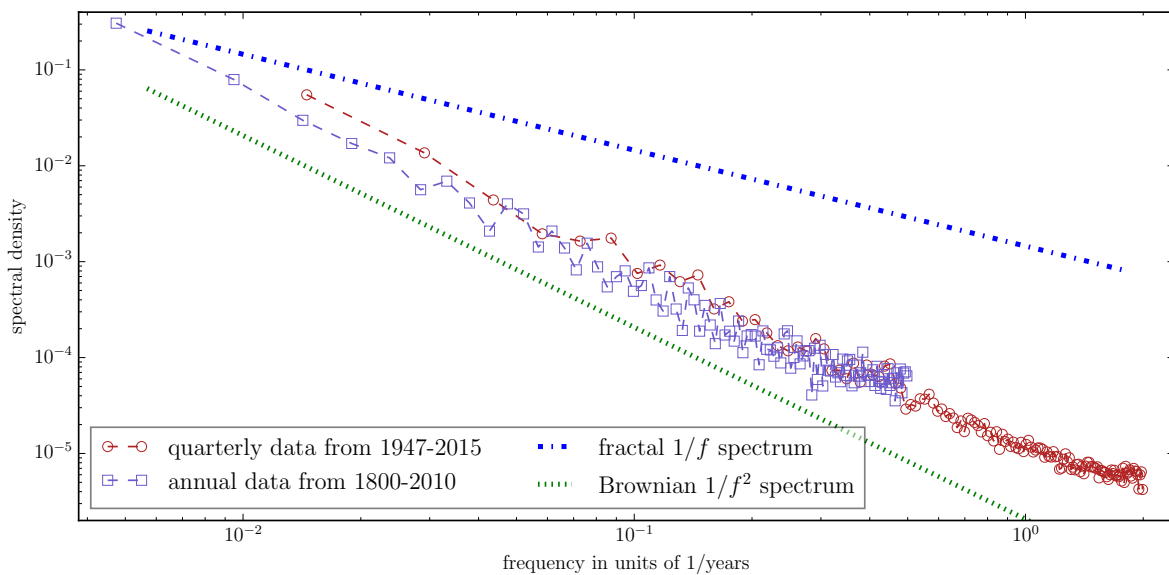


Figure 6.2: Spectral density of r-US-GDP-pc data. We observe a scale-free continuum of scales with no clear peaks. A least squares fit determines an exponent of  $\approx -1.80$  for both the quarterly and the annual data set, thus classifying the GDP as a long-memory process.

frequencies. There are no peak selecting a natural frequency. Instead, all frequencies seem to contribute to the overall GDP dynamics according to a spectral weight following a simple inverse power law function  $P(f) \sim 1/f^p$ , where  $p \approx 1.8$  obtained via a least-squares calibration. In Figure 6.2, we show two lines  $P_{RW}(f) \sim 1/f^2$  (spectral density of the random walk) and  $P_{1/f}(f) \sim 1/f$  ( $1/f$ -spectrum) that are often taken as references. This suggests that the fluctuations of the GDP dynamics are close to that of a random walk, albeit with some departure

making it less volatile than a genuine random walk and closer to a stationary (around its long term trend) long-memory process. We stress that this observation, of a smooth continuous spectrum with no special frequency standing out, casts some doubt on the reliability of previous findings on business cycle periods. We are particularly concerned with the procedure consisting in arguing that business cycle periods should be larger than two years, say, to avoid being contaminated by ‘noise’ at short time scales and should be smaller than ten years, say, to avoid the influence of secular trends [113]. To us, this looks dangerous because it is obvious that one or two characteristic periods will appear as dominant under these imposed conditions. From a time signal analysis perspective, this corresponds to distorting the spectral density by aliasing by a window constraint.

Given this first evidence that all frequencies are important and essentially undistinguishable, we turn to the wavelet transform as a convenient tool to disentangle frequencies appearing at all scales and determine their times of occurrence.

### 6.3 The wavelet transform

Originally developed in geophysics to analyze seismic signals [114], the wavelet transform has proven useful for data analysis in a variety of fields such as image processing [115], astrophysics [116], and turbulence [117]. In economics, the wavelet transform has many useful applications [118, 119], especially in the context of business cycles [120–122].

A  $\psi$ -wavelet transform  $W_\psi$  is simply a projection of a signal  $X(\tau)$  onto  $t$ -translated and  $s$ -dilated versions of  $\psi$  [123]:

$$W_\psi[X](s, t) = \int_{-\infty}^{\infty} d\tau \psi(\tau - t; s) X(\tau). \quad (6.2)$$

We call  $s$  the scale and  $t$  the time parameter. The analyzing function  $\psi$ , called the wavelet, has to be a localized function both in time and frequency domain. Depending on the application, the wavelets must be endowed with several additional properties. See for instance ref. [124] for mathematical details. For our purposes, it is important for the wavelet to be properly normalized. Assuming that  $\psi(t; s)$  is approximately zero for values of  $t$  outside the interval  $[-s, s]$ , the wavelet transform has then the following intuitive interpretation:  $W_\psi[X](s, t)$  is the weighted average of  $X$  over the interval  $[t - s, t + s]$ . The wavelet transform can thus be seen as a ‘mathematical

microscope' that resolves local structures of  $X$  (determined by the shape of the wavelet  $\psi(t; s)$ ) at 'position' (time)  $t$  and at a 'magnification' (scale)  $s$ . Denoting by  $*$  the convolution operator, expression (6.2) can also be written compactly as  $W_\psi[X](s, t) = [X(\tau) * \psi(\tau; s)](t)$ , or, for brevity, just  $X * \psi$ .

Replacing  $\psi$  in (6.2) by its  $n$ -th derivative  $\psi^{(n)}$  corresponds to a  $\psi$ -analysis of the  $n$ -th derivative of the time series  $X(t)$  (up to a normalization factor), as a simple integration by parts derivation shows. In this context,  $\psi = \psi^{(0)}$  is also called the mother wavelet. Since the overall statistical characterization of complex structures depends only weakly on the choice of the mother wavelet [125], we present here only results for the Gaussian mother wavelet  $\psi(t; s) = \exp(-t^2/2s^2)/\sqrt{2\pi}s$ . We have checked that other real-valued mother wavelets give similar results.

In this chapter, we use the wavelet transform to quantify the pattern of local slopes (giving the local growth rates) of the analyzed time series (logarithm of the real US GDP per capita). This amounts to replacing  $\psi$  in (6.2) by the first derivative  $\psi^{(1)}$  of the Gaussian mother wavelet, up to a normalization. The normalization is chosen such that the wavelet transform of the test signal  $X(t) = pt$  with constant slope  $p$  gives exactly its slope  $p$  for all times  $t$  and all scales  $s$ . This leads to the following expression for our analyzing mother wavelet used in expression in (6.2):

$$\psi^{(1)}(t; s) = \frac{t}{\sqrt{2\pi}s^3} \exp\left(-\frac{1}{2}\left(\frac{t}{s}\right)^2\right). \quad (6.3)$$

Note also that, by construction, the wavelet transform performed with  $\psi^{(1)}(t; s)$  of a constant signal is zero, meaning that the wavelet transform is insensitive to the absolute level and only quantifies precisely the local slope at a scale  $s$ .

Our approach uses the wavelet transform to achieve two goals in one stroke: (i) estimate the growth rates and (ii) analyse them at multiple scales. This is different from the approach of Aguiar-Conraria and Soares [121], who first calculated the growth rates by taking the first difference of the logarithm of GDP, and then applied the wavelet transform with a  $\psi^{(0)}$  wavelet to the obtained time series of growth rates. Our integrated approach is better suited to minimize aliasing and biases.

In the remainder of this chapter, all figures are the result of the wavelet transform  $X * \psi^{(1)}$  with  $\psi^{(1)}$  given by (6.3). We focus on the quarterly dataset r-US-GDP-pc dataset, and present a similar analysis for the annual data over the much larger extended period between 1800 and 2010 in the next chapter.

## 6.4 Wavelet analysis of the growth of real US GDP per capita

Plotting the r-US-GDP-pc in a semi-logarithmic plot (Figure 6.3) shows, to a first approximation, a remarkably straight line, suggesting that the GDP grows exponentially as  $\exp(\rho_{lt}t)$  with  $t$  in units of years and a long-term annual growth rate  $\rho_{lt} \approx 2\%$  determined by an ordinary least squares (OLS) fit. This value is often reported in the literature as the average long-term historical growth of real GDP per capita [126].

Beyond this long term average growth, one can see deviations that occur again and again. Moreover, it is interesting to observe that the long-term growth rate  $\rho_{lt}$  represented by the slope of the straight dashed line seems to almost never describe the actual local growth rate of the r-US-GDP-pc. In other words, the average growth rate does not seem to be a good description of the typical growth rates. To quantify these qualitative observations, we perform a wavelet transform analysis of the logarithm of the the r-US-GDP-pc at different times  $t$  and different scales  $s$  to obtain the local growth rate at time  $t$ , averaged over a time interval  $[t - s, t + s]$ , defined by

$$\rho(s, t) = \ln(\text{r-US-GDP-pc}) * \psi^{(1)} . \quad (6.4)$$

The results are encoded with the color scale for the annualized growth rates in Figure 6.3 over the period from 1947 to 2015 shown on the horizontal axis. The left vertical axis plots the scale  $s$  of the wavelet analysis, corresponding to an interval of analysis approximately equal to  $2s$ . For scales at and lower than  $s \approx 4$  years (i.e. averaged over approximately 8 years), one can first observe a hierarchy of branches associated with alternating warm (low or negative growth rates) and cold (positive and strong growth rates) colors. As one goes to smaller and smaller time scales, more fine structures of alternating colors (growth rates) can be seen. At the larger scales,  $s \geq 4$  years, the color settles to the green value, recovering the known, and also directly determined by OLS, long term growth  $\rho_{lt} \approx 2\%$ .

Because the continuous wavelet transform (6.2) contains a lot of redundant information (a function  $X(t)$  of one variable  $t$  is transformed into a function  $W_\psi[X](s, t)$  of two variables  $s$  and  $t$ ), it is standard to compress the wavelet map shown in Figure 6.3 into a so-called ‘skeleton’. The skeleton of  $W_\psi[X](s, t)$  is the set of all local maxima and minima of  $W_\psi[X](s, t)$  considered as a function of  $t$ , for fixed scale  $s$ . The skeleton forms a set of connected curves in the time-scale space, called the extrema lines. Geometrically, each such skeleton line corresponds to either a crest or valley bottom of the three-dimensional representation of the wavelet function  $W_\psi[X](s, t)$ .



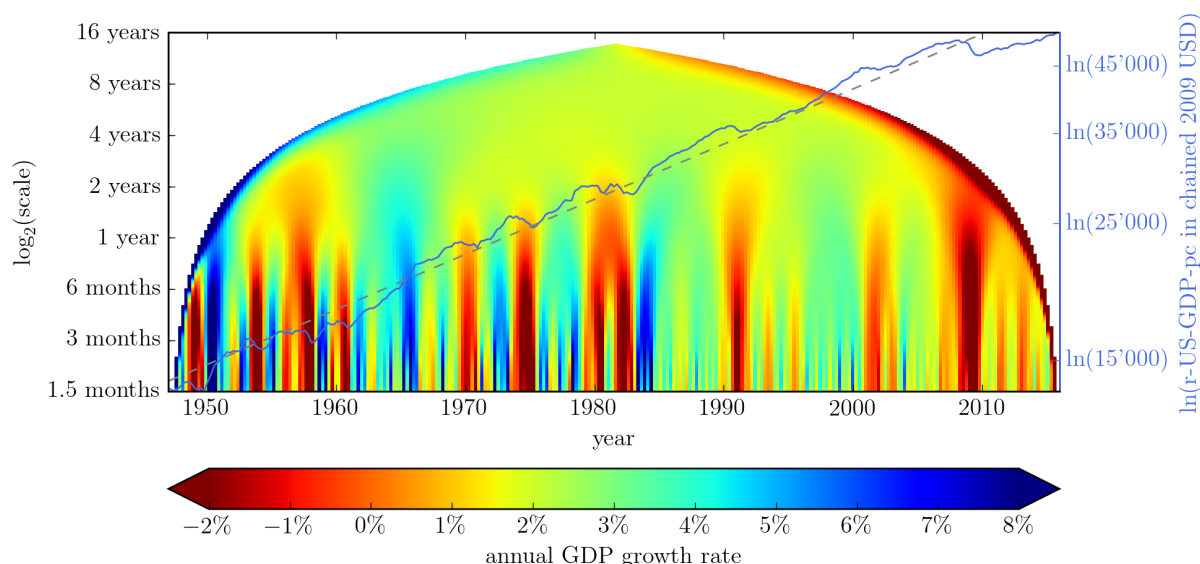


Figure 6.3: Wavelet transform  $\ln(r\text{-US-GDP-pc}) * \psi^{(1)}$  of the logarithm of the quarterly real US GDP per capita data measured in chained 2009 US dollar over the period from 1947 to 2015 and represented by the continuous dark line (right vertical axis). An ordinary least squares fit determines a long-term annualized growth rate  $\rho_{lt}$  of approximately 2%, shown as the dashed line. The left vertical axis plots the scale  $s$  of the wavelet analysis, corresponding to an interval of analysis approximately equal to  $2s$ . The color scale encodes the value of the local annualized growth rates at different times and scales. The nonlinear conical shape of the envelop is due to edge-effects in the wavelet transform. The quarterly growth rates have been rescaled to annual values.

A crest can be viewed as the typical value of the growth rate of a locally surging r-US-GDP-pc. The bottom of a valley is similarly the typical value of the growth rate of a locally slowing down or contracting r-US-GDP-pc. The skeleton structure thus serves as a proxy of (half-) business cycles.

As is clearly visible in Figure 6.4, business cycles are emerging at all scales. There is thus no such thing as ‘the’ business cycle, but rather a continuous hierarchy of overlaid business cycle fluctuations. One can observe clearly the hierarchy of alternating growth regimes, which combine into an overall growth of  $\approx 2\%$  at large scales. Written along each skeleton line in the Figure, we give the values of the local annualized growth rates at four scale levels, 3 months, 6 months, 18 months and 3 years. The structure of the skeleton lines, their colors and the values of the local annualized growth rates confirm the existence of ubiquitous shifting regimes of slow and strong growths. In economics, it has been popular to detrend a time series for its long term growth rate. Our results make this detrending appear somewhat arbitrary, as there is no discrete distinction between short-, medium- and long-term growth.

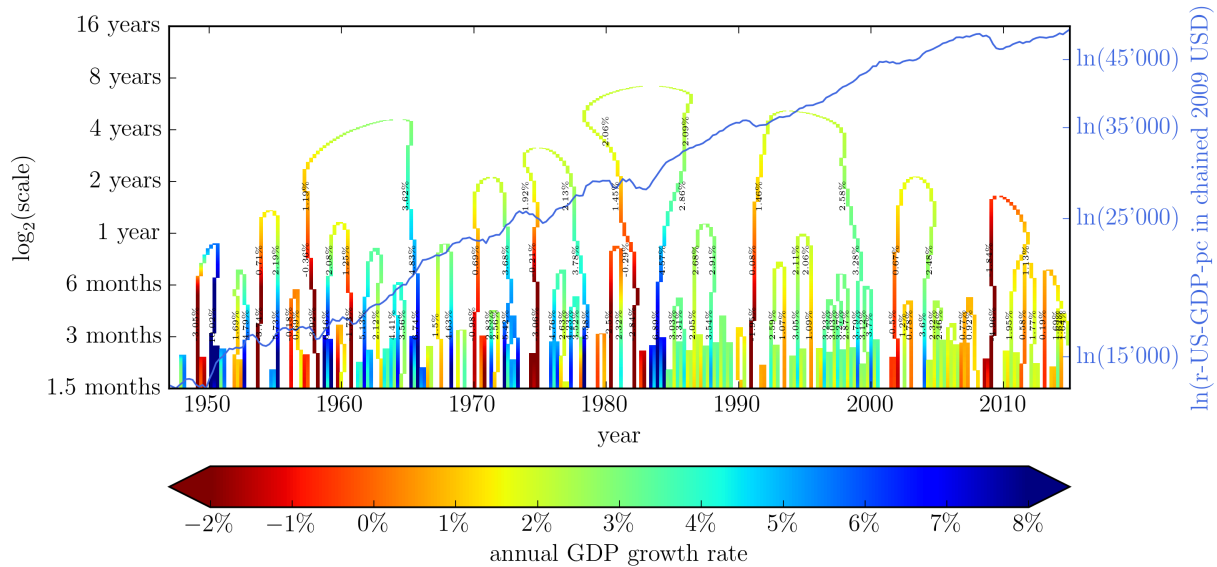


Figure 6.4: Skeleton structure of the wavelet transform  $\ln(r\text{-GDP-pc}) * \psi^{(1)}$  for quarterly real US GDP per capita data measured in chained 2009 US dollar corresponding to Figure 6.3.

## 6.5 Evidence for a robust bimodal structure of distributions of US GDP growth rates

The nature of the shifting regimes of slow and strong growths can be quantified further by constructing the probability density distributions (pdf) of annualized GDP growth rates at different fixed scales, both from the entire wavelet transform (Figure 6.3) and from the skeleton structure (Figure 6.4). The obtained pdf's for four different scales (6, 9, 15 and 30 months) are depicted in Figure 6.5. They have been obtained using a Gaussian kernel estimations with width equal to 0.002. We have checked the robustness of these pdf's by changing the width of the kernels within a factor of two.

The pdf's extracted from the wavelet transform shown in Figure 6.3 and from the skeleton values shown in Figure 6.4 exhibit the same structures. First, the pdf's at the largest scale of 30 months peak at the annualized growth rate of  $\approx 2\%$ , recovering the OLS value reported above (shown as the dashed line in Figure 6.3). Second, as we go down to smaller scales, already at the scale of 15 months, and more pronounced at the scale of 9 and 6 months, a clear bimodal structure emerges (decorated by higher frequency structures, associated with the width of the estimating kernel). Denoting the two main peaks of the bimodal density extracted from the full wavelet transform (the skeleton gives similar results) at scale  $s$  by  $\rho_{\text{low}}(s)$  and  $\rho_{\text{high}}(s)$  respectively, we

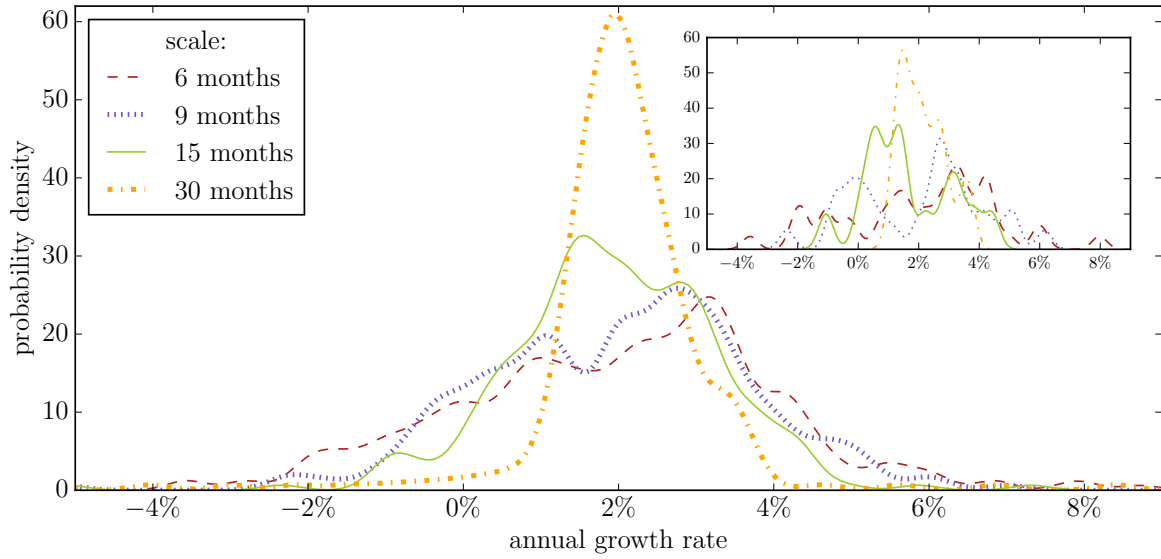


Figure 6.5: Gaussian kernel estimations (with width equal to 0.002) of the probability density distributions (pdf) of the local annualized growth rates of r-US-GDP-pc at four different scales indicated in the inset in the top-left. The main panel represents the distributions extracted from the wavelet transform shown in Figure 6.3, while the top-right inset shows the pdf's obtained from the skeleton values shown in Figure 6.4.

obtain

$$\rho_{\text{low}}(6 \text{ months}) \approx 1\% \lesssim \rho_{\text{low}}(9 \text{ months}) \approx 1.1\% \lesssim \rho_{\text{low}}(15 \text{ months}) \approx 1.5\% \lesssim \rho_{\text{lt}} \approx 2\% \quad (6.5)$$

and

$$\rho_{\text{high}}(6 \text{ months}) \approx 3.1\% \gtrsim \rho_{\text{high}}(9 \text{ months}) \approx 2.8\% \approx \rho_{\text{high}}(15 \text{ months}) \approx 2.8\% \gtrsim \rho_{\text{lt}} \approx 2\%. \quad (6.6)$$

The pleasant stability for the estimates  $\rho_{\text{low}}(6 \text{ months}) \approx \rho_{\text{low}}(9 \text{ months})$  and  $\rho_{\text{high}}(6 \text{ months}) \approx \rho_{\text{high}}(9 \text{ months}) \approx \rho_{\text{high}}(15 \text{ months})$  suggests that real US GDP per capita can be modelled as an alternation of slow growth around a typical value of 1% and strong growth around a typical value of 3%, which bracket the long-term average growth rate  $\rho_{\text{lt}} \approx 2\%$ . This bimodality and the appearance of business cycles at all scales constitute the main results of our chapter.

Finally, we examine the significance of our results from a statistical point of view. How reliable is the observed bimodality, considering that there are only 270 datapoints? How likely is it to observe a bimodal structure, when sampling 270 datapoints from a unimodal density? What is the probability of observing a tri-or multi-modal structure, when sampling from a bimodal density? These questions are formally addressed in the statistical test developed by

|       | 6 months scale | 9 months scale | 15 months scale | 30 months scale |
|-------|----------------|----------------|-----------------|-----------------|
| $H_1$ | 0.05           | 0.16           | 0.22            | 0.94            |
| $H_2$ | 0.31           | 0.62           | 0.47            | 0.09            |

Table 6.1:  $p$ -values for Silverman test, applied to quarterly growth rate distributions at different scales.

Silverman [127]. This allows us to test the null hypothesis,  $H_k$ , that the density underlying the data has  $k$  modes, against the alternative that the density has more than  $k$  modes. Here, we test the quarterly growth rates obtained from the wavelet analysis for both unimodality ( $k = 1$ ) and bimodality ( $k = 2$ ). The  $p$ -value associated with  $H_k$  gives the probability of observing such a sample distribution, assuming that underlying density is truly  $k$ -modal. Large  $p$ -values for test  $H_k$  indicate thus strong statistical evidence for  $k$ -modality. The  $p$ -values, summarized in table 6.1, are in good agreement with our expectations. At smaller scales, the bimodality hypothesis  $H_2$  is clearly preferred over a unimodal distribution. At larger scales, this relationship tips over in favor for the unimodal hypothesis  $H_1$ .

In conclusion, we have presented evidence that the real US GDP per capita growth rate distribution exhibits a bimodal structure. This is, to the best of our knowledge, the first time that the bimodal character of this distribution has been analyzed explicitly. We stress again the subtle, but very important distinction between the interpretation of business cycles when the GDP distribution is unimodal versus bimodal. A unimodal distribution peaked at 2% represents the case where a 2% growth is the norm, in the sense that the bulk of probability mass is concentrated there. In contrast, a bimodal distribution represents the case where a 2% growth rate is uncommon, in the sense of small probability mass in that regime.

In the next chapter we present the wavelet transform and distributions of annual r-US-GDP-pc data starting in 1800 till 2010. The important conclusion is that the previous observations presented above for quarterly data from 1950 to 2015 are broadly confirmed when using annual data over this much longer period.

## 6.6 Conclusions

We have presented a quantitative characterisation of the fluctuations of the annualized growth rate of the real US GDP per capita growth at many scales, using a wavelet transform analysis of two data sets, quarterly data from 1947 to 2015 and annual data from 1800 to 2010. We stress that our use of the chosen mother wavelet (first derivative of the Gaussian function) applied to

the logarithm of the real US GDP per capita provides a robust estimation of the instantaneous growth rate at different scales. Our main finding is that business cycles appear at all scales and the distribution of GDP growth rates can be well approximated by a bimodal function associated to a series of switches between regimes of strong growth rate  $\rho_{\text{high}}$  and regimes of low growth rate  $\rho_{\text{low}}$ . The succession of alternations of these two regimes compounds to produce a remarkably stable long term average real annualized growth rate of 1.6% from 1800 to 2010 and  $\approx 2.0\%$  since 1950.

We thus infer that the robust constant growth rate since 1950 cannot be taken as evidence for a purely exogenous ‘natural’ rate of innovations and productivity growth. It is rather a remarkable output of the succession of booms and corrections that punctuate the economic history of the US since more than 200 years. Our results suggest that alternating growth regimes are intrinsic to the dynamics of the US economy and appear at all scales. These alternating regimes can be identified as generalized business cycles, occurring at the scale of the whole economy.

Such business cycles may be briefly rationalised as follows. During the high growth regime, a number of positive feedback loops are in operation, such as deregulation, enhanced credit creation, the belief in a ‘new economy’ and so on. This creates a transient boom, perhaps accelerating itself and leading to financial and social bubbles [102, 103, 105, 128–130]. This overheating of the economy then turns out not to be sustainable and leads to a correction and consolidation phase, the low growth regime. Then, the next strong growth regime starts and so on.

Our findings suggest that strong growth cannot be dissociated from periods of recessions, corrections or plateaus, which serve as a consolidation phase before the next boom starts. However, because of the remarkable recurrence of the strong regime and in view of its short-term beneficial effects, economists and policy makers tend to form expectations of strong continuous growth. Such way of thinking may lead to conclusions that, we argue, may have little merit. Consider the estimation of the US Federal Reserve Bank of Dallas [131] that the cost of the 2008 crisis, assuming output eventually returning to its pre-crisis trend path, is an output loss of \$6 trillion to \$14 trillion US dollars. These enormous numbers are based on the integration of the difference between the extrapolation of hypothetical GDP trajectories expected from a typical return to pre-crisis growth compared with the realised GDP per capita. In the light of our findings, we argue that it is incorrect to extrapolate to the pre-crisis growth rate, which is by construction abnormally high, and much higher than the long term growth rate. In addition, one should

take into account the fact that the base rate after a crisis should be low or even negative, for the consolidation to work. Moreover, the duration of the boom years may have direct impact on that of the recovery period. In this vein, the 2008 crisis is special, as it is the culmination of a 30 year trend of accelerating financialization, deregulation and debt growth [101]. Our present results impel the reader to ponder what is the ‘natural’ growth rate and avoid naive extrapolations.

Using a simple generic agent-based model of growth [100], the existence of a trade off between either low and steady growth or large growth associated with high volatility has been identified. Translating this insight to the US economy and combining with the reported empirical evidence, the observed growth features shown in the present chapter seem to reveal a remarkable stable relationship between growth and its fluctuations over many decades, if not centuries. Perhaps, this is anchored in the political institutions as well as in the psychology of policy makers and business leaders over the long term that transcend the short-term vagaries of political power sharing and geopolitics. It is however important to include in these considerations the fact that the US is unique compared with other developed countries, having benefitted enormously from the two world wars in particular (compared with the destruction of the French, Japanese and UK empires and the demise of the economic dominance of European powers).

## Chapter 7

# GDP growth rates as confined Lévy flights: towards a unifying macro theory of economic growth rate fluctuations <sup>†</sup>

In chapters 2-5 we have focused on drift-free random walks constrained by repulsive barriers, as observed in forex target zones. In this chapter, we will consider a different, more granular time series: the (US) gross rate domestic product. We will provide evidence that GDP growth rates can be modeled as a specific type of random walk confined in steep potential walls.

As a result of the increasing interconnectedness of firms, and more general, of society [132], the study of firm size fluctuations and its effect on the entire economy has become an active area of research [133–138]. A central question is how the productivity fluctuations of individual firms aggregate into productivity measures for the economy as a whole. Albeit still subject to debate [138], there is now a dominant view [133–137], that idiosyncratic shocks to individual large firms can have a significant effect on the macroeconomic output growth. Explanations for this effect range from statistical fluctuation arguments [133] to more intricate models taking into consideration intersectoral input-output linkages [134] or dynamic income-expenditure networks [137].

In this chapter, we offer a different, more coarse-grained perspective. We show how heavy-

---

<sup>†</sup>This chapter is based on reference [14].

tailed idiosyncratic shocks are mitigated through a kind of renormalized economic potential to the aggregate output growth distribution. Specifically, we claim that the aggregate output growth rate distribution can be modeled as a Lévy flight in a steep confining potential. This establishes a quantitative connection between the tail distribution of idiosyncratic shocks to individual firms and the aggregate economic growth distribution. To provide empirical support for our claim, we study the gross domestic product (GDP) of the United States over that last 200 years, and find that the model is in good agreement with the data.

## 7.1 Structure of the GDP growth rate distribution

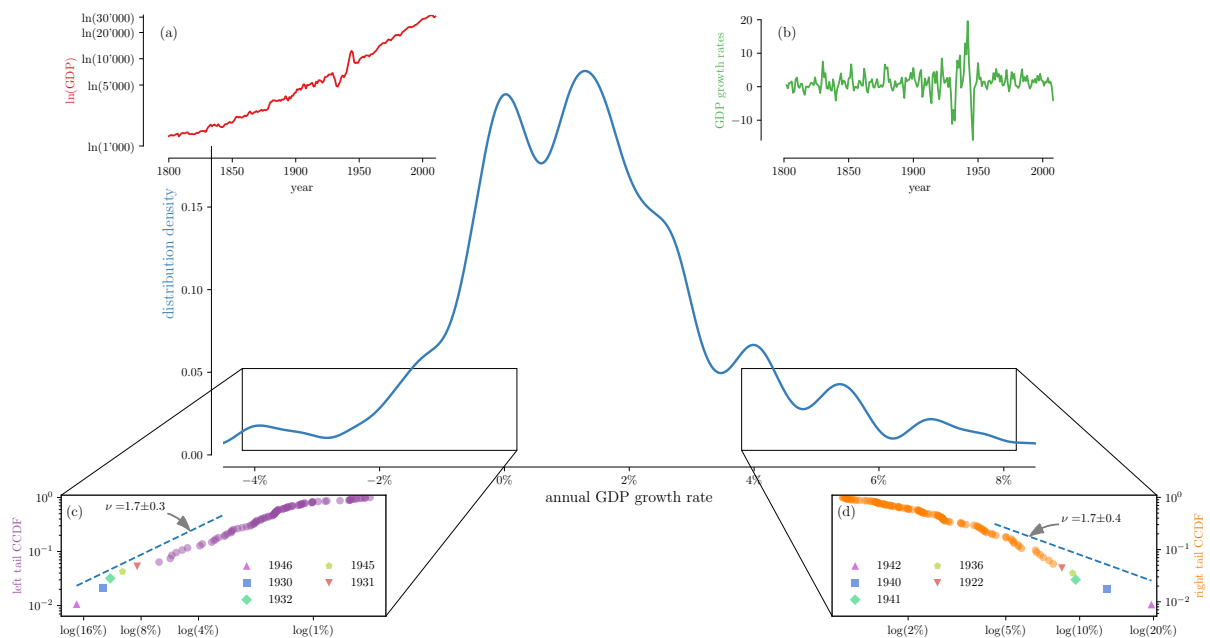


Figure 7.1: Top left figure (a) shows the real US GDP per capita over the last 200 years. Using a wavelet transform, we extract the roughly 200 annual growth rates shown in the top right figure (b). The growth rate distribution is then obtained using a Gaussian kernel density estimation, shown in the main figure. The bimodal structure is readily visible in the middle of the distribution. The left- and right- tails of the distribution are shown in a double-logarithmic plots in figures (c) and (d), respectively. The right tail is measured in absolute values, whereas the left tail is measured in terms of absolute percentage deviation from 1.5%, to avoid negative arguments. We observe symmetric tails with CCDF tail exponents roughly between 1.5 and 2. This is robust with respect to removal of the largest positive and negative growth rates, associated with the two world wars, the great depression and the impact on 1922 of the sharp deflationary recession of 1920-21 following the end of the WWI.

As a measure of overall economic growth, we analyze the real US GDP per capita (from hereon, simply GDP) over the past 200 years (Figure 7.1(a)). In contrast, the unnormalized total nominal GDP contains growth contributions as a result of population growth and inflation,



both of which are not true sources of increased output productivity per individual. Given the scrutiny of the GDP in economic research, the distribution of its growth rates has also received a lot of attention. In this section, we discuss two important characteristics of that distribution, its heavy tails and its bimodality.

### 7.1.1 Evidence of bimodality

In the previous chapter, we have discussed in depth the bimodality of the US GDP growth rate distribution and its implications. Here, we present a similar analysis, but instead of using quarterly data over the past 50 years, we use annual data over the past 200 years. We will see that also here the distribution is bimodal, and its two peaks are rationalized by the fundamental out-of-equilibrium nature of the economy that is switching between boom and bust states, i.e. business cycles. To extract the annual GDP growth rates, we use a smoothing wavelet transform. A  $\psi$ -wavelet transform  $W_\psi$  of a time-series  $X(\tau)$  is just given by the convolution

$$W_\psi[X](s, t) = \int_{-\infty}^{\infty} d\tau \psi(\tau - t; s) X(\tau). \quad (7.1)$$

Given that the wavelet  $\psi$  is approximately zero outside the interval  $[-s, s]$ , the wavelet transform is interpreted as the weighted average of  $X(t)$  over the interval  $[t - s, t + s]$ . Using integration by parts, we note that replacing  $\psi$  in (7.1) by its  $n$ -th derivative  $\psi^{(n)}$  corresponds to a  $\psi$ -analysis of the  $n$ -th derivative of the time-series  $X$  (up to a scaling factor). Here, we are interested in growth rate changes, and hence we work with the first derivative of a Gaussian wavelet. The normalization is chosen such that the wavelet transform of a test signal  $X(t) = mt$  with constant slope  $m$  gives exactly its slope  $m$  at all times  $t$  and scales  $s$ , specifically,  $\psi^{(1)} = \frac{t}{\sqrt{2\pi}s^3} \exp\left(-\frac{1}{2}\left(\frac{t}{s}\right)^2\right)$ . Applying the wavelet transform to the log-GDP data (Figure 7.1(a)) at a one year scale results in the time-series of GDP growth rates depicted in Figure 7.1(b). Using a kernel density estimator, we can then extract the growth rate distribution, shown in the main plot of Figure 7.1. Here, we have used a Gaussian kernel with a bandwidth of 0.3% such that, to the eye, the bimodal structure in the middle of the distribution is apparent. But one has to be careful, as this bimodal appearance is susceptible to the choice of bandwidth. If it is too small, the individual datapoints give the impression of spurious multimodality. If too large, the bimodal pattern is washed out. To obtain statistically sound evidence for this bimodal structure, we apply Silverman's bimodality test [127]. Specifically, we test for the null hypothesis,  $H_k$ , that the density underlying the data

has  $k$  modes, against the alternative that it has more than  $k$  modes. The associated  $p$ -values read  $p_1 = 6\%$ ,  $p_2 = 19\%$  and  $p_3 = 5\%$ , clearly in support of the claim that the density is bimodal.

### 7.1.2 Evidence of heavy tails

It is now a broadly accepted stylized fact that the distribution of GDP growth rates has tails that are fatter than Gaussian [139, 140], and can even be power-law [141, 142] (but see also ref. [143] for a counter argument).

Figures 7.1(c) and (d) depict complementary cumulative distribution functions (CCDF, also called survival function) of the left and right tail of the distribution, respectively. The five largest positive and negative growth rates, associated with the two world wars and the great depression, are annotated explicitly. By applying likelihood ratio tests of the power-law against the family of stretched-exponentials [144], we find  $p$ -values of 96% and 61% for left- and right-tail respectively. We conclude that both tails are parsimoniously represented in terms of pure power laws, without the need for the more general parameterization of the stretched exponential distribution family. Using standard procedures [145], we then estimate the CCDF tail exponent  $\nu$  and its estimated standard error  $\sigma$  as  $\nu = n [\sum_{i=1}^n \log(r_i/r_{\min})]^{-1}$  and  $\sigma = \nu/\sqrt{n}$ , with  $r_1, \dots, r_n$  the (absolute value of)  $n$  growth rates larger than some  $r_{\min}$  above (below) which the right (left) power-law tail is active. The exponents of left and right tail are surprisingly symmetric, and robust with respect to removal of the smallest and largest growth rates,  $r_{\min}$ , or the Gaussian kernel bandwidth, leading merely to slightly increased error intervals. The tail exponent  $\nu = 1.7$  being smaller than 2 qualifies the tails as belonging to the Lévy-stable distribution regime [146].

In line with previous findings [13, 139–142], we have shown that the GDP growth rate distribution exhibits a bimodal structure with symmetric power-law tails. In the next section, we present a model that relates these two seemingly independent properties.

## 7.2 Lévy flights in steep confining potentials

A generic representation of the Markovian dynamics of the GDP growth rate in continuous time is given by

$$dr = -\frac{dV}{dr} dt + dL(\mu, D) \quad (7.2)$$

where  $r = r_t$  is the growth rate at time  $t$  and  $V$  is some function, called the ‘potential’, which ensures a nonlinear mean-reversing dynamics. The negative derivative of the potential is then

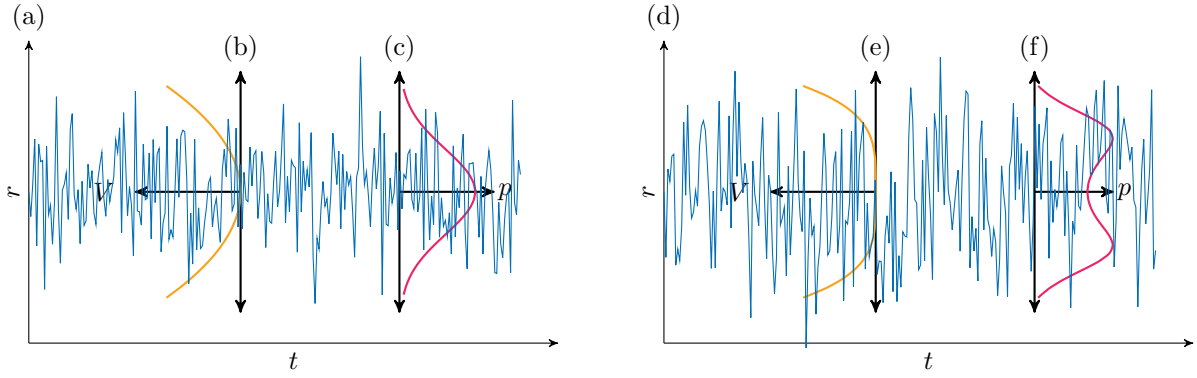


Figure 7.2: Figure (a) shows the realization of an Ornstein-Uhlenbeck process (7.2) with  $\beta = 2, \mu = 2, D = 1$ . Figure (b) depicts the quadratic potential  $V(r) = \alpha r^2/2$  it is generated from and figure (c) is the asymptotic unimodal density distribution. Figure (d) shows the realization of a process (7.2) in a steep potential with Lévy-stable distributed noise, with parameters  $\beta = 4, \mu = 1.5, D = 1$ . Figure (e) depicts the corresponding quartic potential  $V(r) = \alpha r^4/4$  and figure (f) is the asymptotic bimodal density distribution.

interpreted as the ‘force’ that is guiding the growth rate. The stochastic contribution  $dL(\mu, D)$  denotes Lévy-stable distributed noise with tail exponent  $\mu \in (0, 2]$  and scale-parameter  $D > 0$ .

The most common mean-reverting model of type (7.2) is the Ornstein-Uhlenbeck process, characterized by a square potential ( $V(r) = \alpha r^2/2, \alpha > 0$ ) and Gaussian noise ( $\mu = 2$ ). With basic stochastic methods [147], it can be shown that, for the Ornstein-Uhlenbeck process, the asymptotic stationary distribution of  $r_t$  is Gaussian with standard deviation  $D/\sqrt{2\alpha}$ . This is illustrated in Figure 7.2 (left). In presence of Gaussian noise, the stationary distribution remains unimodal even when the square potential is replaced by a steeper potential

$$V(r) = \frac{\alpha}{\beta} |r - r_0|^\beta \quad (\beta > 2). \quad (7.3)$$

Here, we have also introduced a potential midpoint  $r_0$  for more generality. Similarly, the distribution remains unimodal for a square-potential in presence of a heavy-tailed source of noise ( $\mu < 2$ ). However, counter-intuitively, it has been shown [148, 149] that the combination of a steep potential (7.3), with heavy-tailed noise, results in a bimodal stationary distribution, as sketched in Figure 7.2 (right). Furthermore, the asymptotic left and right tail of the distribution are power-laws with CCDF tail exponent  $\nu \equiv \beta + \mu - 2$ . In particular, this implies that, for  $\beta > 4 - \mu$ , the steep potential walls confine the Lévy noise to the extent that its tails have finite variance.

### 7.3 Fitting the Lévy flight parameters

Before applying the Lévy flight to GDP data in the next section, we discuss here how to fit the parameters using maximum likelihood, and test the fitting accuracy on a synthetic dataset. Process (7.2) with potential (7.3) depends on five parameters  $\alpha, \beta, r_0, \mu$  and  $D$ . Since the tail exponent  $\nu$  of the growth rate distribution can be measured independently, we could, in principle, use the constraint  $\nu = \mu + \beta - 2$  to fit either  $\mu$  or  $\beta$  and then deduce the other one. But as we demonstrate now, fitting all five parameters at the same time works remarkably well, and so we can use the constraint imposed on  $\nu$  to validate the theory.

Consider a set of  $n + 1$  sequential datapoints (e.g. GDP growth rates)  $r_1, r_2, \dots, r_{n+1}$ . We derive the  $n$  growth rate increments  $\Delta r_i = r_{i+1} - r_i$ , and assume an equal time step of  $\Delta t$  between two observations (e.g. 1 year). The discretized version of (7.2) is given by [150]

$$\Delta r_i = \alpha \operatorname{sign}(r_0 - r_i) |r_i - r_0|^{\beta-1} \Delta t + L(\mu, D) (\Delta t)^{1/\mu} \quad (7.4)$$

where  $L(\mu, D)$  is a realization a Lévy stable random variable with parameters  $\mu$  and  $D$ . Rearranging terms yields

$$L(\mu, D) = \frac{\Delta r_i + \alpha \operatorname{sign}(r_i - r_0) |r_i - r_0|^{\beta-1} \Delta t}{(\Delta t)^{1/\mu}}. \quad (7.5)$$

In other words, the datapoints  $\{z_i\}_{i=1}^n$ , with

$$z_i \equiv \frac{\Delta r_i + \alpha \operatorname{sign}(r_i - r_0) |r_i - r_0|^{\beta-1} \Delta t}{(\Delta t)^{1/\mu}}, \quad (7.6)$$

are  $L(\mu, D)$  distributed. Denote by  $\phi(\cdot; \mu, D)$  the density distribution of the random variable  $L(\mu, D)$ . We can then determine the parameters  $\alpha, \beta, r_0, \mu$  and  $D$  by maximizing (the logarithm of) the likelihood function

$$\mathcal{L}(\alpha, \beta, r_0, \mu, D) = \prod_{i=1}^n \phi(z_i; \mu, D). \quad (7.7)$$

Note that the dependence on the parameters  $\alpha, \beta$  and  $r_0$  is only implicit through the datapoints  $z_i$ . To test the fitting accuracy, we have generated synthetic datapoints according to (7.4). To stay close to the application following in the next section, we have generated 200 datapoints with exactly the parameters  $\alpha, \beta, r_0, \mu$  and  $D$  that we determine from the fit to 200 datapoints of annual GDP growth rates below. The likelihood profile plots in Figure 7.3 have clear maxima

close the the true parameter values for all five parameters, thus indicating that good convergence can be expected. To confirm this intuition, we have used again the same parameters to simulate 1'000 times 200 datapoints from (7.4), and fitted the parameters for each realization. The resulting distributions of the parameter estimates, depicted in Figure 7.3, confirm that the obtained fits are very reliable. In the next section, we use the bootstrapped 80%-confidence interval as error bars.

## 7.4 Growth rate fluctuations as confined Lévy flights

We propose that the Lévy flight in a steep potential is a good model for economic growth rates. Concretely, we see the Lévy noise as a representation of the firm or sector specific idiosyncratic shocks. Indeed, firm size growth rates are known to be well approximated by heavy-tail distributions with infinite variance [133]. A priori, it is not clear that fluctuations of individual firms (or sectors) would have a significant impact on aggregate measures for the economy. According to the central limit theorem, the average fluctuations of the sum of the outputs of  $n$  firms decay as  $\sim 1/\sqrt{n}$ , a negligible effect in the limit of large  $n$ . The state of the economy is then merely influenced by economy-wide shocks such as oil price shocks, currency devaluation, wars, etc. However, such arguments break down if the distribution of firm sizes is very heavy tailed [133], or the interdependencies of firms are asymmetric [134]. In a model that takes the microscopic firm structure into account, the individual firm shocks are propagated across the economy through a network of input-output linkages. To what extent the individual shocks are averaged out upon aggregation depends on the topology of the network of firm and sector dependencies [134]. If a small number of firms or sectors play a disproportionately important role as input suppliers to others, the economy is more susceptible to fluctuations of these few firms, in contrast to a more balanced scenario. Empirical evidence for the US [134], France [135] and Sweden [136] suggests that indeed such asymmetries in the inter-firm and sector networks lead to significant contributions of a few large companies or sectors to the overall economic performance of a country.

In our model, the microscopic firm network is represented by an effective ‘economic potential function’. The advantage of this coarse-grained perspective is that it does not rely on any specific assumptions about the underlying generating mechanisms at the micro level. The details average out and are captured in the potential parameter  $\beta$ . The steeper the potential, the more the

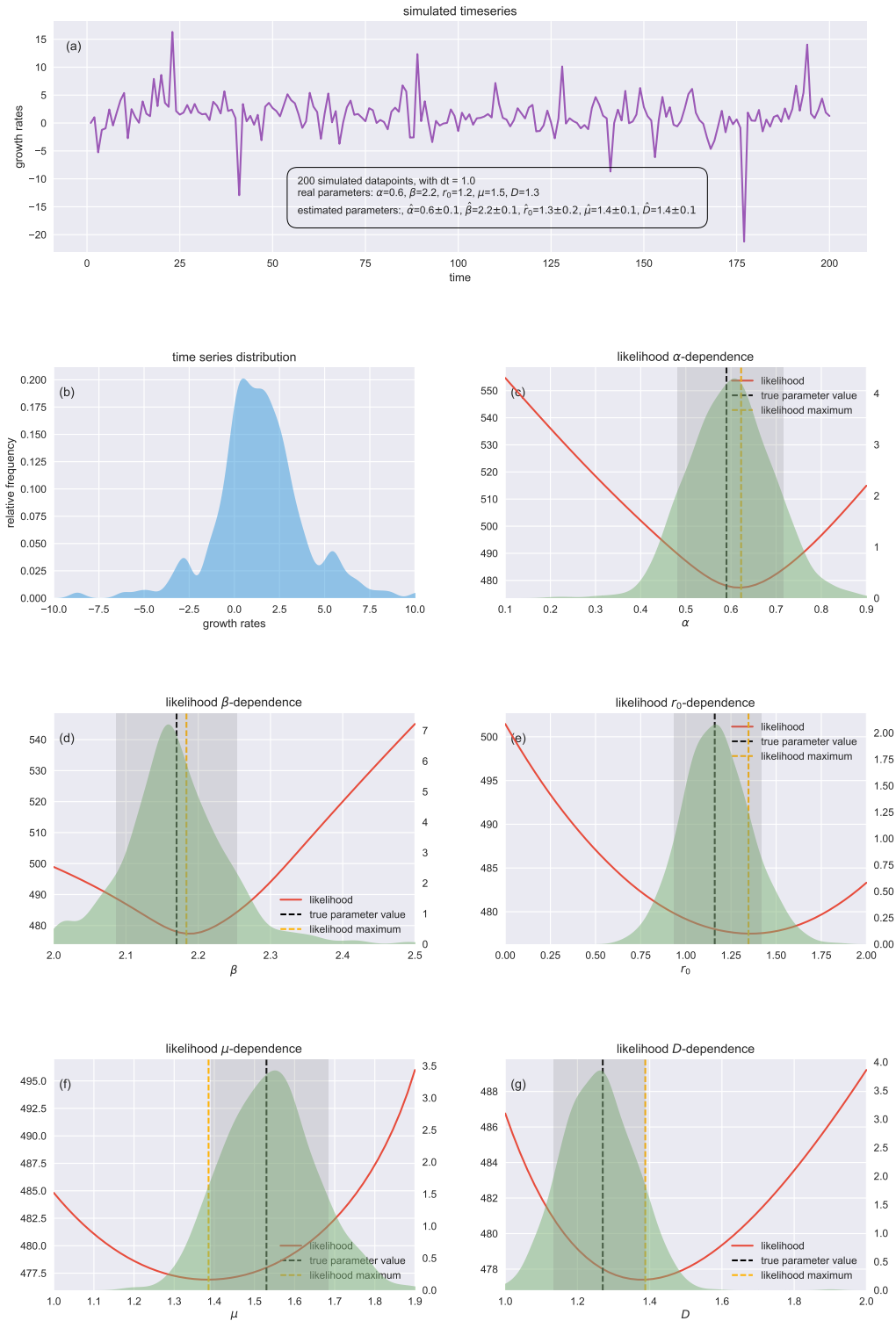


Figure 7.3: Given the fitted parameters from from section 7.4 (Figure 7.4), we simulate 200 yearly datapoints according to (7.4). Figure (a) shows the realized time-series and figure (b) the corresponding distribution density. The left axes of figures (c)-(g) depict (the negative logarithm of) the likelihood (7.7) as a function of the parameter  $\alpha$ ,  $\beta$ ,  $r_0$ ,  $\mu$  and  $D$ , respectively, while the other parameters are held fixed to their true values. It can be seen that the profile minima are reasonably close to the true values for all five parameters. To quantify this, we have simulated 1'000 such Lévy flights and estimated the parameters. On the right axes of figures (c)-(g), we show the resulting distribution of the parameter estimates. The grey bands depict the empirical 80%-confidence intervals that we use as error bars.

individual firm shocks average out, and the less heavy the aggregate tail. The restitution force is thus acting to damp the idiosyncratic shocks. This approach is motivated by the renormalisation group approach [151] that provides the template to represent the collective effect of many degrees of freedom at the micro level by a few effective ‘renormalised’ degrees of freedom at the macro level.

For more quantitative support of our proposition, we now fit the roughly 200 growth rate datapoints to the model, just as explained in the previous section. The outcome is visualized in Figure 7.4. We find that the fitted parameters are robust with respect to removal of individual data-points, with the exception of the two, in absolute value, largest growth rates, associated with World War II. To obtain statistically robust estimates, we thus remove these two largest datapoints. Beyond statistical arguments, this is justified conceptually in the sense that a world war is a global shock that affects the economy as a whole. Since we are interested in contributions of individual shocks to aggregate economic output measures, it is justified to remove such events. Technically, we should thus get rid of all economy-wide shocks. But since only the aggregate GDP growth rates are analyzed, this is technically infeasible, and not necessary insofar as our calibration results are not sensitive to individual datapoints, except for the WWII case. WWII was indeed special for the U.S, which became the factory of the world, doubling its GDP from 1940 to 1943. If we were to keep the WWII datapoints, the potential force changes slightly to  $\alpha = 0.8$  and  $\beta = 2.1$ , whereas the other parameters remain unaffected.

The predicted GDP tail exponent  $\nu = \beta + \mu - 2 = 1.7 \pm 0.2$  is in excellent agreement with the actual tail exponent found equal to  $1.7 \pm 0.3$  (Figure 7.1(c,d)). The magnitude of the microscopic noise with  $\mu = 1.5 \pm 0.2$  is lighter than Cauchy noise ( $\mu = 1$ ), but still clearly separated from normality ( $\mu = 2$ ). The potential, with an exponent of  $\beta = 2.2 \pm 0.1$ , is not much steeper than a quadratic potential with the linear mean-reverting force (and even closer to a quadratic potential if we do not omit the WWII growth rates). It is just sufficiently steeper than quadratic to ensure a binomial structure of the distribution of growth rates in the presence of Lévy-noise but close enough to the quadratic case so that the GDP growth rate distribution has diverging second moment, in line with other findings [141, 142]. The potential midpoint  $r_0 = 1.2 \pm 0.3$  is more centered towards the left of the two bimodal peaks. However, as the simulation in Figure 7.5 confirms, this is well within the realms of expectations when analyzing just 200 datapoints.

As a final cross check, we compare the autocorrelation  $\langle r_t r_{t-1} \rangle$  of the simulation and the real GDP data. As a correlation measure, we use Kendall’s  $\tau$  instead of the more common Pearson

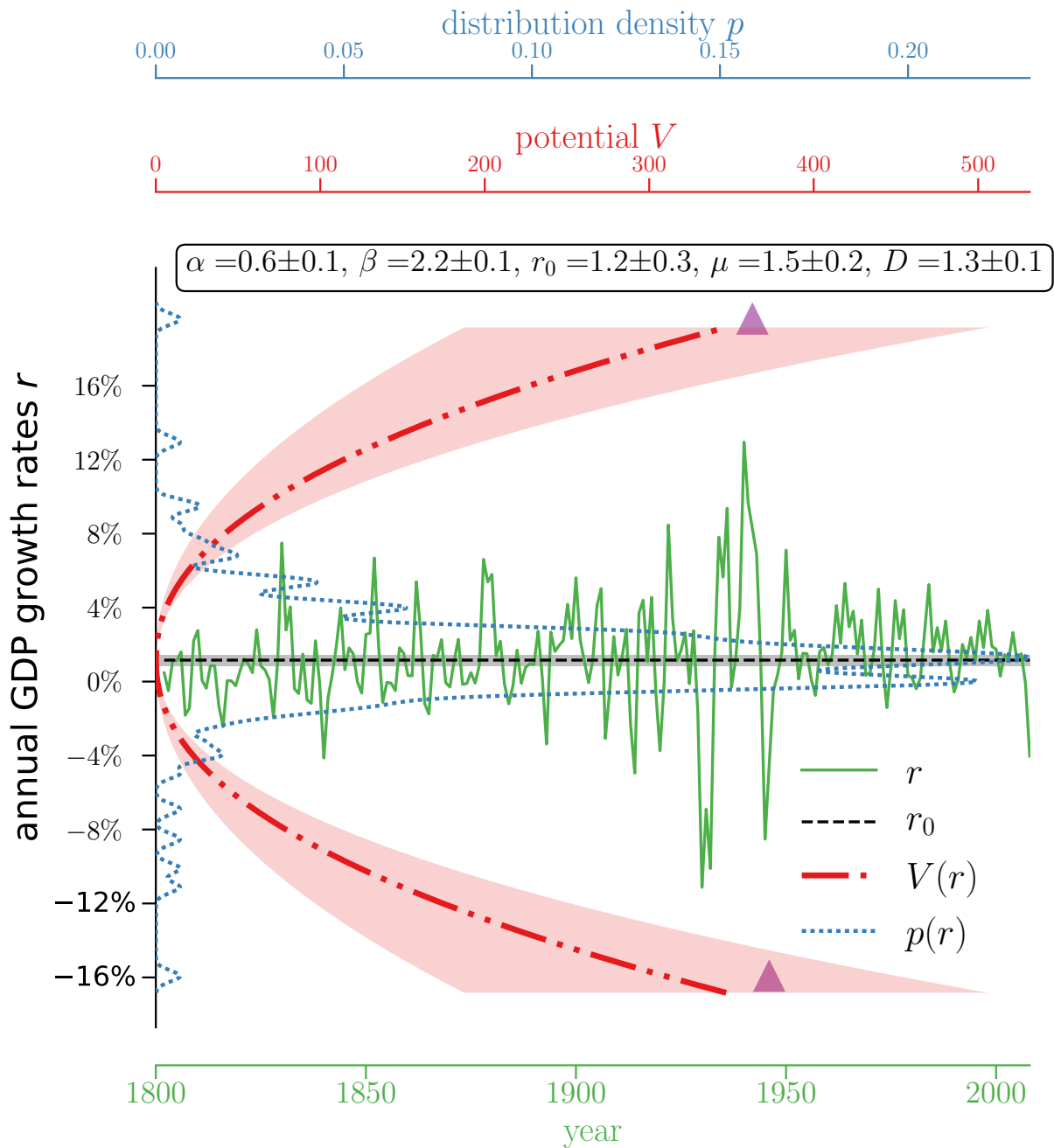


Figure 7.4: Using maximum likelihood, we have fitted the Lévy flight model (7.2) in the steep potential (7.3) to annual GDP growth rates, but omitting the two largest growth rates in absolute value, associated with WWII. We find  $\alpha = 0.6 \pm 0.1$ ,  $\beta = 2.2 \pm 0.1$ ,  $r_0 = 1.2 \pm 0.3$ ,  $\mu = 1.5 \pm 0.2$  and  $D = 1.3 \pm 0.1$ . The green line shows the annual GDP growth rates, as obtained from the wavelet analysis (same as Figure 7.1(b) but without the largest WWII datapoints). The red dotted line (and its surrounding error band) denotes the aggregate potential force (7.3) with the fitted values for  $\alpha$ ,  $\beta$  and  $r_0$ . The black dashed line (and its surrounding error band) annotates the potential center-point  $r_0$ . The blue dotted line shows the GDP growth rate distribution (same as main plot in Figure 7.1). The two purple triangles annotate the two in absolute value largest GDP growth rates, associated with WWII (c.f. Figure 7.1(c,d)) that we have discarded from the fit, as explained in the text.



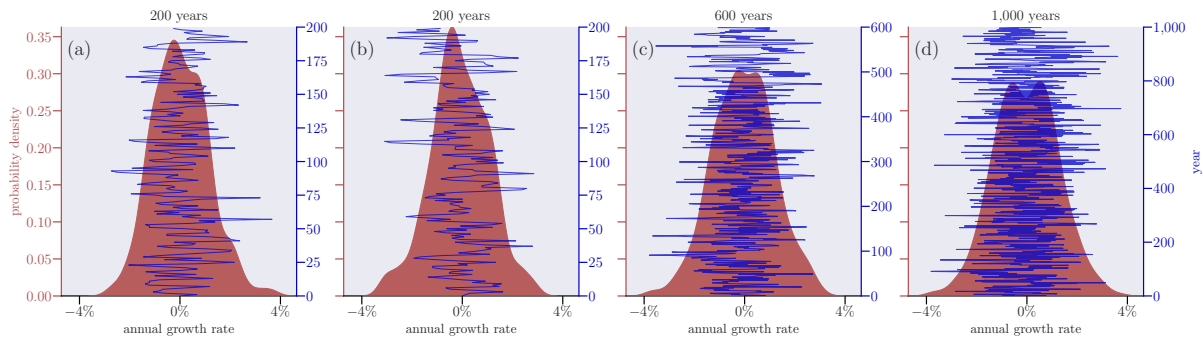


Figure 7.5: With the same parameters as in Figure 7.3 (but  $r_0 = 0$ ), we have simulated two times 200, 1'000 and 2'000 yearly datapoints of Lévy flights. Figures (a) and (b) have been simulated using a different random number seed. The bimodal structure is more evident in the first of the two realizations. We have used the same seed for figures (b), (c) and (d). As we increase the number of samples, the bimodal structure becomes more pronounced.

correlation, to avoid convergences issues related to the heavy tails of the process. Using again the fitted parameters from Figure 7.4, we have simulated 1'000 Lévy flights over 200 years. The average autocorrelation that we measure is  $0.18 \pm 0.05$  where we have used the standard deviation of the 1'000 flights as error bars. Empirically, we measure a correlation of 0.23 for the actual GDP data, which is just within the error bars of the simulation.

## 7.5 Conclusions

In conclusion, we have shown that GDP growth rates are well described as Lévy flights in a potential that is steeper than quadratic. With only minimal ingredients, this model is able to capture the aggregate effect of idiosyncratic shocks to averaged economic output growth measures. It thereby establishes a connection between the tail exponents at a micro and macro scale, while simultaneously accounting for the bimodal structure of business cycle fluctuations. These are promising results that draw novel connections, and we hope that our findings will spark future research.



## Chapter 8

# Quantification of the evolution of firm size distributions due to mergers and acquisitions <sup>†</sup>

In chapters 2-5, we have described exchange rates as free random walks constrained by fixed boundaries. In chapters 6 and 7, we have described GDP growth rates as Lévy flights embedded in steep potential walls, i.e. ‘soft boundaries’. Both models have focused on the description of an individual time-series subject to one or several constraints. Here, in contrast, we consider the evolution of multiple individual stochastic processes (firm sizes), subject to occasional ‘interaction constraints’ in the form of merger and acquisitions.

In any established economy, companies of all sizes can be observed, ranging from small local businesses up to multinational corporations. The size of a company can conveniently be measured for instance in terms of yearly revenue, asset value, number of employees, and so on. It is by now a well-established fact that the distributions of firm sizes are heavy tailed [152, 153]. Knowledge of the firm size distribution is important because it has direct implications for an economy, for instance for monetary policy [154], employment [155] or innovation [156]. Generally, a heavy tailed firm size distribution implies that a large fraction of market impact is due to a few large firms that dominate the economy [133, 157]. It is then not surprising that the study of firm size distributions has received a great deal of attention both in the theoretical and empirical literature.

While theoretical models for the heavy tail nature of firm size distributions are manifold, at

---

<sup>†</sup>This chapter is based on reference [15].

the heart of most explanations is the idea of proportional growth (Gibrat's law [158]). Gibrat's law states that the growth rate of a company is independent of its absolute size, or, equivalently, that the change in size is linearly proportional to its size. In absence of any other ingredient, the resulting firm size distribution is a log-normal function. It is noteworthy that, when the variance (in the log-variable) of the log-normal function is sufficiently large, it can appear close to a power law in an intermediate range of sizes [159–161]. Empirically however, the distribution of firms is fatter tailed than a log-normal function, and thus requires other ingredients than just proportional growth.

In addition to proportional growth, the various mathematical models have been enriched with other realistic ingredients such as firm heterogeneity [162–165], minimum firm size [166], births and deaths of firms [167–169], and sudden bankruptcy [170, 171]. Zipf's law is then usually obtained due to subtle balance conditions between the growth rate and available external resources [166, 171, 172]. Finally, deviations from Gibrat's fundamental assumption of proportional growth have also been considered [170, 173]. Sutton [174] gives a chronological wrap-up of the study of firm size distributions and a detailed overview on recent empirical studies is provided by Segarra and Teruel [175].

The theoretical models underlying the papers referenced above focus mainly on growth through investments in a company's own operations (internal growth). The large impact of mergers and acquisitions (external growth) is most often not taken into account explicitly. For instance, over the period of 1995-1999, Schenk [176] observes that investments in acquisitions by North American and West European firms were approximately equal to sixty per cent of gross investments in machinery and equipment and they easily outpaced those in Research and Development (R&D). Investments in acquisitions were no less than about eight times higher than business enterprise expenditures on R&D. Although it must be mentioned that the period of 1995-1999 falls in the midst of a so-called merger wave [177, 178], these facts nevertheless make the point that mergers and acquisitions (M&A) are likely to be highly relevant for the study of firm size distributions. Taking into considerations M&A (as well as its counter-part, the spin-offs) constitutes an important component of a holistic understanding of the distribution of firm sizes.

In this chapter, we take a first step in this direction by formalizing the effect of M&A on the firm size distribution in terms of a non-linear partial integro-differential equation. This equation is formally equivalent to the coagulation equation from physics. By isolating the effect

of M&A on the firm size distribution, we show that the coagulation equation allows to account for various empirical observations. Furthermore, our approach is very flexible and can conveniently be extended to also take into consideration spin-offs, internal growth, firm birth, bankruptcy and other growth related phenomena.

The remainder of this chapter is structured as follows. In section 8.1, we summarize the literature that has studied the effect of mergers and acquisitions on firm size distributions. In section 8.2 and 8.3, we suggest a quantitative explanation for these observations in terms of the coagulation equation. Section 8.4 establishes a connection between the empirical findings and our theoretical calculations that let us account for the empirical stylized facts. Section 8.5 confirms further our approach by comparison with numerical solutions. Section 8.6 concludes and discusses possible extensions of the model.

## 8.1 The effect of mergers and acquisitions on Zipf's law

While there is a large amount of literature examining Zipf's law in general, there are only a few studies that investigate its relation to mergers and acquisitions. In this section, we extract the main conclusions from those studies formulated in terms of four key observations 1-4.

1. Examining a sample of large American firms, Ijiri & Simon [179] notice that smaller firms have a higher chance of being absorbed. Singh [180] concludes the same when studying firm sizes in the UK. Aaronovitch and Sawyer [181] confirm this finding by observing that UK firm sizes and probability of acquisitions are inversely related. We will see in the next section how this property has a very simple mathematical expression in terms of the coagulation equation.
2. Ijiri & Simon [182] examine the firm size distributions in 1956 and 1957 in a sample of large American firms. They conclude that the distribution of the 500 largest firms remained relatively unchanged from the effect of M&A. Their analysis thus supports the proposition that firm growth due to M&As would follow Gibrat's law of proportional growth to the same extent as internal growth.
3. In a later study, Ijiri & Simon [179] examine the 831 largest US industrial firms in 1969 and conclude that the firm size distribution is clearly affected by M&A. Plotting the firm size distribution in a log-log plot after the M&A had taken place, they find that the distribution

with M&As is shifted slightly to the right of the distribution without M&As. On top of that, both distributions are concave and not straight lines expected under Zipf's law. In section 8.4, we will see how the difference between Ijiri & Simon's study [179, 182] finds a natural explanation in terms of the coagulation equation.

4. Cefis et al. [183] examine the entire population of Dutch manufacturing firms. Unlike previous studies, they examine an entire population of firms and not only large firms. They conclude that M&As do not affect the size distribution when they consider the entire population of firms. Examining only firms that are at some point involved in an M&A, a shift of the firm size distribution towards larger sizes is noticed. This shift is not uniform but affects firms of different sizes in different ways. While the number of firms in the lower tail decreased, the number of firms in the central size classes increased.

A fifth observation is by Hannah and Kay [184]. Studying UK manufacturing firms in 1957, the authors decompose the growth rate into internal and external contributions and conclude that the observed power law coefficient is attributed to external growth. Furthermore, had it been just for internal growth, smaller firms would have grown faster than larger firms. These findings were later challenged and debated [185, 186]. This fifth observation addresses the interplay between internal and external growth. Albeit interesting and debated, accounting for this observation is not part of the current work and will not be further mentioned. However, as will become clear below, our model can be extended to also capture the mixed dynamics of external and internal growth.

In the subsequent sections, we set up a model that allows us to account for observations 1-4.

## 8.2 Mergers and acquisitions as a coagulation process

Let us denote by  $p(m, t)$  the distribution of firm sizes  $m$  at time  $t$ . We propose to model the effect of M&A on  $p$  as an equation that consists of two dynamic contributions: On the one hand, the distribution density at size  $m$  is increased due to the merger of a firm of size  $m'$  with a firm of size  $m - m'$  to a firm of size  $m$ . On the other hand, the density at size  $m$  is decreased through the merger of a company of size  $m$  with a company of size  $m'$  to a company of size  $m + m'$ . Since this logic applies to all firm sizes  $m'$ , we have to take the integral ("sum") over all such

contributions, giving rise to the following integro-differential equation

$$\frac{dp(m, t)}{dt} = \frac{1}{2} \int_0^m dm' A(m', m - m') p(m', t) p(m - m', t) - p(m, t) \int_0^\infty dm' A(m, m') p(m', t). \quad (8.1)$$

We call  $A(m, m')$  the merger kernel, describing the rate at which a firm of size  $m$  merges with a firm of size  $m'$  to produce a firm of size  $m + m'$  (we neglect here any dilution of size that may occur during the merger). The first term on the right-hand side of (8.1) accounts for all the firms of size  $m'$  that merge at time  $t$  with a company of size  $m - m'$  to form a company of size  $m$ . Mergers of this type are weighted with the rate  $A(m', m - m')$  and then the ‘sum’ over all such  $m', m - m'$  pairs is taken. The factor  $1/2$  is to avoid double counting. The second term on the right hand side decreases the concentration of firms of size  $m$  by subtracting the sum of all firms of size  $m$  that merge at time  $t$  with a firm of size  $m'$  to form a firm of size  $m + m'$ . We choose here a continuous description for mathematical convenience. Alternatively, the integrals in (8.1) can be replaced by discrete sums. This does not qualitatively alter the presented results.

As was pointed already out by Saichev et al. [170], equation (8.1) is well-known in physics as Smulowski’s coagulation equation [187, 188]. This equation has been introduced to describe the evolution of the number density of particles with mass or size  $m$  as they consolidate (coagulation process). The coagulation equation has found many applications in astrophysics, cloud physics, polymer chemistry and formation of aerosols [189–191]. See Figure 8.1 for an illustration of this concept.

Considering the huge amount of literature in physics and mathematics that has been devoted to equation (8.1), it should be convenient to fall back on one of the many known solutions. It turns out, however, that our merger equation is different from what is usually considered in physical systems. This is for two reasons. First, as is discussed in the next section, our merger kernel  $A(m, m')$  is a decaying function in  $m$  and  $m'$ . Most physical systems assume that the coagulation kernel  $A(m, m')$  be a (homogeneous and) increasing function in  $m$  and  $m'$ . Second, at time  $t = 0$ , we start from a heavy tailed firm size distribution. In physical applications, one usually starts from a thin tailed or even point-like initial distribution. This changes significantly the solution methods discussed in the literature. Only recently, power law initial distributions in combination with decaying coagulation kernels have been considered [192–195]. However, merely approximate scaling results in the limits of small and large masses are known and several issues are not yet clearly resolved. The state of the art from a mathematical perspective is well

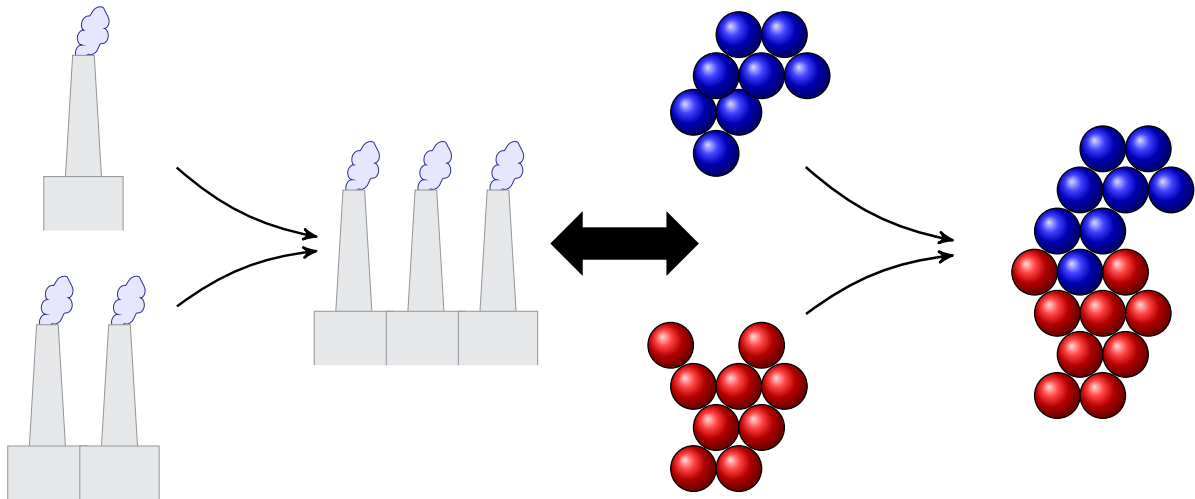


Figure 8.1: Both the merger and acquisition of firms (left side) and the coagulation of physical particles (right side) are formally described by the same mathematical equation (8.1). The fact that the two processes are described by the same mathematical equation just reflects the generic process of growth by assembling.

summarized by Costa [196]. We conclude that there are so far no known solutions to the merger equation that we develop and solve in the next section.

The coagulation equation (8.1) is essentially just one of many potential growth mechanisms, which can be summarized in the following population-balance equation:

$$\frac{\partial p(m, t)}{\partial t} = \left[ \frac{\partial p(m, t)}{\partial t} \right]_{\text{M\&A}} + \left[ \frac{\partial p(m, t)}{\partial t} \right]_{\text{spin-off}} + \left[ \frac{\partial p(m, t)}{\partial t} \right]_{\text{int. growth}} + \left[ \frac{\partial p(m, t)}{\partial t} \right]_{\text{exogenous}}. \quad (8.2)$$

The M&A component is given by (8.1). The spin-off off dynamics obeys

$$\left[ \frac{\partial p(m, t)}{\partial t} \right]_{\text{spin-off}} = 2 \int_0^{\infty} d\tilde{s} K(m, \tilde{m}) p(m + \tilde{m}) - p(m, t) \int_0^m d\tilde{m} K(m - \tilde{m}, \tilde{m}) \quad (8.3)$$

with  $K$  the spin-off kernel. The interpretation of (8.3) is very similar to the one of (8.1). The first term on the right hand side of (8.3) captures the split of a firm of size  $m + \tilde{m}$  into two firms of size  $m$  and  $\tilde{m}$ , respectively. The second term accounts for all the firms of size  $m$  that split into two smaller firms. Internal growth can be captured for instance through the Fokker-Planck equation

$$\left[ \frac{\partial p(m, t)}{\partial t} \right]_{\text{int. growth}} = -\frac{\partial}{\partial m} [\mu(m) p(m, t)] + \frac{1}{2} \frac{\partial^2}{\partial m^2} [\sigma(m)^2 p(m, t)] \quad (8.4)$$

with the functions  $\mu$  and  $\sigma$  proportional to  $m$ , according to Gibrat's law of proportional growth [158]. The exogenous growth term in (8.2) captures the various additional influences such as



sudden bankruptcy, creation of new firms etc.

Equation (8.2) can, in general, only be solved numerically, for which many parameters have to be fixed (see however Chap. 10 in [170], which offers an analytical solution in the case of non-proportional growth in the presence of M&A and spin-offs). Instead, here we isolate the perturbative effect of M&A on the firm size distribution, for which we can obtain an approximate analytical solution. This allows us to relate our results to previous empirical studies. We account (partially) for the existence of additional growth mechanisms by starting from an initial Zipf's law distribution, that can be generated under various different conditions of internal growth, firm birth and firm bankruptcy [171]. This assumption is justified insofar as M&As have become especially prominent in the second half of the last century (cf. Figure 1 in [178]), whereas the observations of heavy tailed firm size distributions date back to Gibrat [158]. We thus study the perturbative effect of M&A on an already developed firm size distribution.

Finally, we can neglect spin-off activity, which occurs with a much smaller frequency than M&A's, as illustrated by the Japanese firm database provided by the corporate research company TDB. For 2014, for instance, out of a total of roughly 1,100,000 firms, there was approximately 4,500 mergers and acquisitions and less than 100 spin-off events.

### 8.3 Coagulation equation for heavy tailed distributions: an approximate solution

The solution  $p(m, t)$  of (8.1) describes the firm size distribution under the influence of M&A at time  $t$ . We assume that our observation starts at time  $t = 0$  from a heavy tailed initial distribution. Here, time  $t$  is an abstract parameter that will be related to real time only in the next section. In this section, we give an analytical first order solution to (8.1).

#### 8.3.1 Specification of the merger kernel and initial distribution

The coagulation equation (8.1) is not yet fully determined. We must also specify a merger kernel  $A(m, m')$  and an initial distribution  $p(m, t = 0)$ . The kernel  $A(m, m')$  describes the rate at which a firm of size  $m$  merges with a firm of size  $m'$  to form a firm of size  $m + m'$ . According to observation 1, this must be a decaying function in  $m$  and  $m'$ . A natural choice is thus

$$A(m, m') = e^{-\alpha m} + e^{-\alpha m'} \quad (8.5)$$

for some  $\alpha > 0$ . Clearly, other kernels are possible. In particular, our intuition tells us that a take-over becomes more likely with the acquiring firms becoming relatively more dominant in size. We will see below that there is no strong dependence of our results on the kernel.

As stated in the introduction, the exact shape of the firm size distributions is debated. We could thus choose either a log-normal or a Pareto as initial distribution. Since the log-normal is considerably more difficult to treat analytically, we decide to work here with a power law. Numerical results for the log-normal are found in section 8.5. Assuming a power law initial distribution means  $p(m, t = 0) = \mu m_0^\mu / m^{1+\mu}$  for some lower cut-off  $m_0$  and power-law exponent  $\mu$ . Zipf's law is recovered exactly by setting  $\mu = 1$ . On the other hand, as already mentioned, the log-normal law with sufficiently large variance (in the log-variable) resembles a power law with  $\mu \gtrsim 0$  over a wide range of values [159–161]. Empirically, various values of  $\mu$  have been reported, ranging from  $\mu = 1/2$  up to values close to  $\mu = 2$ . Segarra and Teruel [175] provide a good overview.

Working with a power-law  $p(m, t = 0) = \mu m_0^\mu / m^{1+\mu}$  brings along two inconveniences. First, its Laplace transform involves the incomplete Gamma function that is analytically difficult to handle. Second, we must introduce an arbitrary lower cut-off  $m_0$ . A way to avoid both of these problems is to work with the fractional exponential distribution (FED) [197] introduced by as a generalization of a Poisson process with long memory. For  $0 < \mu \leq 1$ , the FED is of the form [198]

$$f_\mu(m) = -\frac{d}{dm} E_\mu(-m^\mu), \quad (8.6)$$

or, equivalently [199],

$$f_\mu(m) = m^{\mu-1} E_{\mu,\mu}(-m^\mu), \quad (8.7)$$

with the Mittag-Leffler functions  $E_\mu$  and  $E_{\mu,\nu}$  defined by

$$E_\mu(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\mu n + 1)} \quad \text{and} \quad E_{\mu,\nu}(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\mu n + \nu)}. \quad (8.8)$$

Repin and Saichev [198] show that  $f_\mu$  is asymptotically a power law:

$$f_\mu(m) \approx \begin{cases} m^{\mu-1} / \Gamma(\mu), & m \rightarrow 0; \\ m^{-(1+\mu)} / \mu \Gamma(1-\mu), & m \rightarrow \infty. \end{cases} \quad (8.9)$$

What makes the FED extremely useful compared to a strict Pareto initial distribution - in fact,

this is its defining property - is its simple Laplace transform,

$$\hat{f}_\mu(k) = \int_0^\infty dm e^{-mk} f_\mu(m) = \frac{1}{1+k^\mu}. \quad (8.10)$$

By working with  $f_\mu$ , we restrict ourselves to exponents  $\mu \in (0, 1]$ . Once again, numerical results in section 8.5 will confirm that our results are not strongly dependent on the exact choice of  $\mu$ .

### 8.3.2 First order analytical solution

There are only a few known analytical solutions to equation (8.1). All of them involve a kernel  $A$  that is an increasing function in  $m$  and  $m'$  and most explicit solutions consider initial distributions with finite support. Here we consider an exponentially decaying kernel and initial distribution  $f_\mu$  with a power law tail. We derive first order analytical results that account for empirical observations already very well.

We start by noting that kernel (8.5) is of the form

$$A(m, m') = F(m) + F(m') \quad (8.11)$$

for some function  $F$  (here  $F(m) = \exp(-\alpha m)$ ), which allows for a convenient simplification [200], as the following calculations show: Plugging (8.11) into the coagulation equation (8.1) gives

$$\begin{aligned} \frac{dp(m, t)}{dt} &= \int_0^m dm' F(m') p(m') p(m - m') - F(m) p(m) M_0(t) - p(m) \int_0^\infty dm' F(m') p(m') \\ &= [Fp * p](m) - F(m) p(m) M_0(t) - p(m) \int_0^\infty dm' F(m') p(m') \end{aligned} \quad (8.12)$$

where we have defined  $M_0(t) \equiv \int_0^\infty dm p(m, t)$  and introduced the convolution operator ‘\*’. Now, we make the substitution

$$q(m, \theta) = \frac{p(m, t)}{M_0(t)}, \quad d\theta = dt M_0(t), \quad (8.13)$$

implying

$$\frac{d}{dt} [p(m, t)] = \frac{d}{dt} [M_0(t)q(m, \theta(t))] \quad (8.14a)$$

$$= q \frac{dM_0(t)}{dt} + M_0 \frac{dq}{dt} \quad (8.14b)$$

$$= q \frac{dM_0(t)}{dt} + M_0 \frac{d\theta}{dt} \frac{dq}{d\theta} \quad (8.14c)$$

$$= q \frac{dM_0(t)}{dt} + M_0^2 \frac{dq}{d\theta}. \quad (8.14d)$$

Plugging this substitution into (8.12) gives

$$q \frac{dM_0}{dt} + M_0^2 \frac{dq}{d\theta} = M_0^2 [Fq * q] - FM_0^2 q - M_0^2 q \int_0^\infty dm' F(m')q(m'). \quad (8.15)$$

Matters simplify once we notice that the first term on the lhs of (8.15) cancels with the third term on the rhs of (8.15). Indeed,

$$q \frac{dM_0}{dt} = q \int_0^\infty dm' \frac{dp(m', t)}{dt} \quad (8.16a)$$

$$= q \int_0^\infty dm' \left[ M_0^2 [Fq * q] - FM_0^2 q - M_0^2 q \int_0^\infty d\tilde{m} F(\tilde{m})q(\tilde{m}) \right]. \quad (8.16b)$$

We note that the first term of (8.16b) cancels with the third term of (8.16b) since

$$\int_0^\infty dm' [Fq * q](m') = \int_0^\infty dm' \int_0^{m'} dn [Fq](n) q(m' - n) \quad (8.17a)$$

$$= \int_0^\infty dm' \int_0^\infty dn [Fq](n) q(m' - n) \quad (8.17b)$$

$$= \int_0^\infty dn [Fq](n) \int_0^\infty dm' q(m' - n) \quad (8.17c)$$

$$= \int_0^\infty dn [Fq](n) \int_0^\infty d\tilde{m} q(\tilde{m}) \quad (8.17d)$$

where we have used in (8.17b) that  $q(m) = 0$  for  $m < 0$  by assumption and Fubini's theorem in (8.17c). Hence,

$$q \frac{dM_0}{dt} = -M_0^2 q \int_0^\infty dm' F(m')q(m') \quad (8.18)$$

which confirms that the first term in (8.15) cancels with the third term on the rhs of (8.15).

In conclusion, substitution (8.13) transforms the coagulation equation (8.1) into the simpler equation

$$\frac{dq(m, \theta)}{d\theta} = [Fq * q](m) - F(m)q(m). \quad (8.19)$$

Taking the Laplace transform of (8.19), we find

$$\frac{d\hat{q}(k, \theta)}{d\theta} = [\hat{q}(k, \theta) - 1] \widehat{Fq}(k, \theta). \quad (8.20)$$

Equation (8.20) holds for general coagulation kernels of the form (8.11), in particular for (8.5). The function  $F(m) = \exp(-\alpha m)$  induces merely a shift by  $\alpha$  in the Laplace transform, so that (8.20) becomes

$$\frac{d\hat{q}(k, \theta)}{d\theta} = [\hat{q}(k, \theta) - 1] \hat{q}(k + \alpha, \theta), \quad (8.21)$$

where  $\hat{q}(k)$  is the Laplace transform of  $q(m)$ . Equation (8.21) is a mixed functional differential equation [201]. Analytical solutions to non-linear equations of this type are not common. Instead, we make use of the fact that  $\alpha$  must be small. The coagulation kernel describes the rate at which a company of size  $m$  merges with a firm of size  $m'$ . Since mergers are observed also between large firms (as measured for instance in terms of revenue or number of employees), kernel (8.5) must assign non-vanishing probability also to events with  $m, m' \gg 1$ . Consequently,  $\alpha \ll 1$  and we can approximate  $\hat{q}(k + \alpha)$  with a first order Taylor expansion,  $\hat{q}(k + \alpha) \approx \hat{q}(k) + \alpha \hat{q}'(k)$  and plug this into (8.21), giving

$$\frac{\partial \hat{q}(k, \theta)}{\partial \theta} + \alpha(1 - \hat{q}) \frac{\partial \hat{q}}{\partial k} = \hat{q}(\hat{q} - 1). \quad (8.22)$$

This quasilinear first-order PDE can be solved with the method of characteristics. We look for a parametrization of the form  $k(s), \theta(s), \hat{q}(k(s), \theta(s))$ , resulting in the three ODEs

$$\frac{d\theta}{ds} = 1, \quad (8.23)$$

$$\frac{dk}{ds} = \alpha(1 - \hat{q}), \quad (8.24)$$

$$\frac{d\hat{q}}{ds} = \hat{q}(\hat{q} - 1), \quad (8.25)$$

with initial conditions

$$\theta(s = 0) = 0, \quad k(s = 0) = k_0, \quad \hat{q}(s = 0) = g(k_0) \quad (8.26)$$

and  $g(\cdot)$  the initial condition. Here,  $k_0$  is only formally defined, as a function that relates  $k$  to  $s$  (see appendix). Equation (8.23) is trivial, giving  $\theta = s$ . We can thus drop the artificial parameter  $s$  and write  $\theta$  instead. Next, we solve equation (8.25) by separation of variables:

$$\frac{d\hat{q}}{\hat{q}(\hat{q}-1)} = d\theta \quad (8.27a)$$

$$\log\left(\frac{1-\hat{q}}{\hat{q}}\right) = \theta + \text{const} \quad (8.27b)$$

$$\hat{q}(\theta) = \frac{1}{1 + \text{const} \cdot e^\theta}. \quad (8.27c)$$

According to (8.26), we require that at  $\theta = 0$ , (8.27c) satisfies

$$g(k_0) \stackrel{!}{=} \frac{1}{1 + \text{const}} \quad (8.28)$$

where  $g$  is our initial distribution. As discussed above  $p(m, t = 0)$  is the fractional exponential  $f_\mu(m)$  with Laplace transform

$$g(k_0) = \hat{f}_\mu(k_0) = \frac{1}{1 + k_0^\mu}. \quad (8.29)$$

Comparing with (8.28) then fixes the integration constant to  $k_0^\mu$  so that the final solution to equation (8.25) reads

$$\hat{q}(k_0, \theta) = \frac{1}{1 + k_0^\mu e^\theta}. \quad (8.30)$$

Of course, this solution holds also for general initial distribution  $g$ . Solving (8.28) for ‘const’ just gives

$$\hat{q}(k_0, \theta) = \frac{1}{1 + \left(\frac{1}{g(k_0)} - 1\right) e^\theta} \quad (8.31)$$

which recovers (8.30) as the special case  $g = f_\mu$ . We plug (8.30) into (8.24),

$$\frac{dk(\theta)}{d\theta} = \alpha \left(1 - \frac{1}{1 + k_0^\mu e^\theta}\right). \quad (8.32)$$

Because the rhs is independent of  $k$ , the solution of (8.32) is merely an integration problem:

$$k = \alpha \log\left(1 + k_0^\mu e^\theta\right) + \text{const}. \quad (8.33)$$

The initial value (8.26) determines the integration constant, and we end up with

$$k(k_0, \theta) = \alpha \log \left( \frac{1 + k_0^\mu e^\theta}{1 + k_0^\mu} \right) + k_0. \quad (8.34)$$

As is explained in the appendix, the final result to (8.22) is therefore

$$\hat{q}(k, \theta) = g(k_0(k, \theta), \theta) = \frac{1}{1 + k_0^\mu(k, \theta)e^\theta}, \quad (8.35)$$

where  $k_0(k, \theta)$  is the inversion of (8.34). Equation (8.34) can only be inverted numerically. However, note that for some fixed  $m$ , the relevant contribution of the Laplace transform  $\hat{q}(k)$  to  $q(m)$  comes from  $k$  values that are in the range  $k \sim \mathcal{O}(\frac{1}{m})$ . It is then straight forward to see from (8.35) that for realistic values of  $\alpha$  and  $\theta$  the approximation  $k \approx k_0$  holds very well for the entire range of relevant  $k$  values (see Figure 8.2). In that case, (8.35) simplifies to

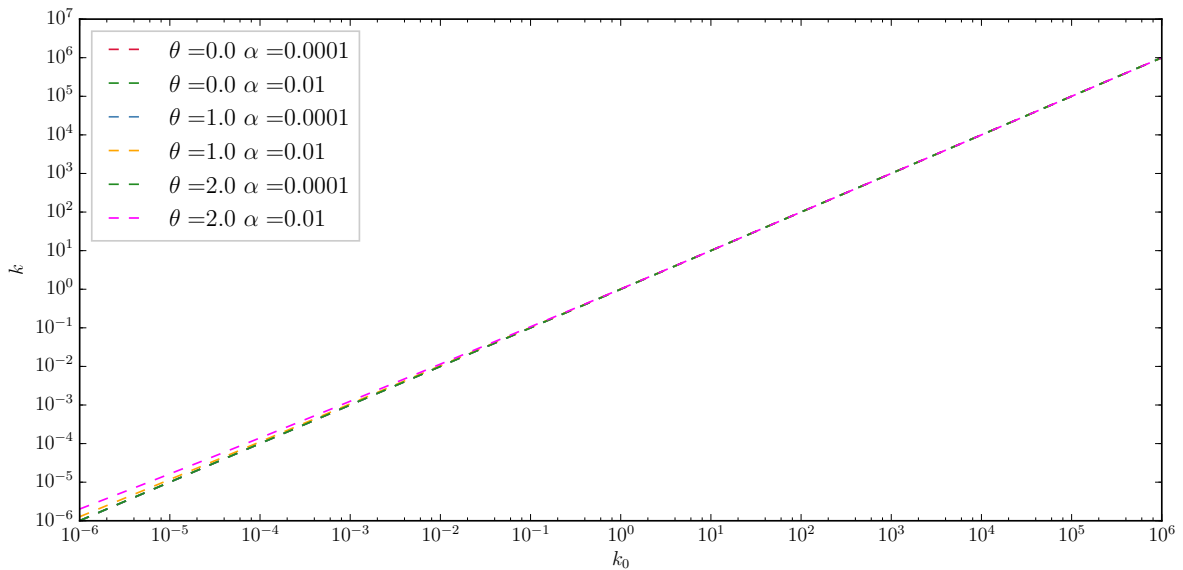


Figure 8.2: We show the functional relation (8.34) between  $k$  and  $k_0$  for different values of  $\alpha$  and  $\theta$ . We can see that the approximation  $k \approx k_0$  holds ver well over the entire range of relevant  $k$  values.

$$\hat{q}(k, \theta) = \frac{1}{1 + k^\mu e^\theta}. \quad (8.36)$$

We notice that (8.36) is almost exactly of the form (8.29), only with a prefactor  $\exp(\theta)$ . Because the inverse Laplace transform is an integration problem, we can relate (8.36) to the inverse

Laplace transform of (8.29) simply by substitution of a constant. The result reads

$$q(m, \theta) = e^{-\theta} m^{\mu-1} E_{\mu, \mu} \left( -e^{-\theta} m^{\mu} \right). \quad (8.37)$$

This finishes the calculation of  $q(m, \theta)$ . Finally, we have to transform back to  $p(m, t)$ . First, we integrate the second equation in (8.13). Using that  $\theta(t=0) = 0$ , this gives

$$\theta = \int_0^t d\tau M_0(\tau). \quad (8.38)$$

We note that  $M_0(t) = \hat{p}(0, t)$ . Hence, by taking the Laplace transform of the coagulation equation (8.1) and then setting  $k = 0$  we arrive at an equation for  $M_0$ . Looking at (8.12), the Laplace transform of (8.1) is straight forward,

$$\frac{d\hat{p}(k, t)}{dt} = \widehat{F}p(k)\hat{p}(k) - M_0\widehat{F}p(k) - \hat{p}(k)\widehat{F}p(0). \quad (8.39)$$

With  $F(m) = \exp(-m\alpha)$ , this translates into

$$\frac{d\hat{p}(k, t)}{dt} = \hat{p}(k + \alpha, t)\hat{p}(k) - M_0\hat{p}(k + \alpha, t) - \hat{p}(k)\hat{p}(\alpha, t) \quad (8.40)$$

and setting  $k = 0$  gives

$$\frac{dM_0}{dt} = -\hat{p}(\alpha, t)M_0 = -\hat{q}(\alpha, \theta(t))M_0^2 = -\frac{M_0^2}{1 + k^{\mu}e^{\theta}}. \quad (8.41)$$

This last equation involves  $\theta$  explicitly via (8.36). To get rid of this dependency, we self-consistently approximate  $\theta(t) = \theta'(0)t + \mathcal{O}(t^2) \approx t$ . This approximation is justified a posteriori below. Plugging  $\theta \approx t$  into (8.41) yields

$$\frac{dM_0}{dt} = -\frac{M_0^2}{1 + \alpha^{\mu}e^t}, \quad M_0(0) = 1. \quad (8.42)$$

This Riccati equation is straight forward to solve with separation of variables. We obtain

$$M_0(t) = \left( 1 + t + \log \left( \frac{1 + \alpha^{\mu}}{1 + \alpha^{\mu}e^t} \right) \right)^{-1} = 1 - \frac{t}{1 + \alpha^{\mu}} + \mathcal{O}(t^2), \quad (8.43)$$



and then, via (8.38),

$$\theta = t - \frac{t^2}{2(1 + \alpha^\mu)} + \mathcal{O}(t^3). \quad (8.44)$$

In conclusion, the first order solution to (8.1) is given by (8.37) with  $\theta(t)$  determined via (8.44).

This solution is depicted in Figure 8.3 for  $\mu = 0.5$  at different times  $t$ . We notice that, for

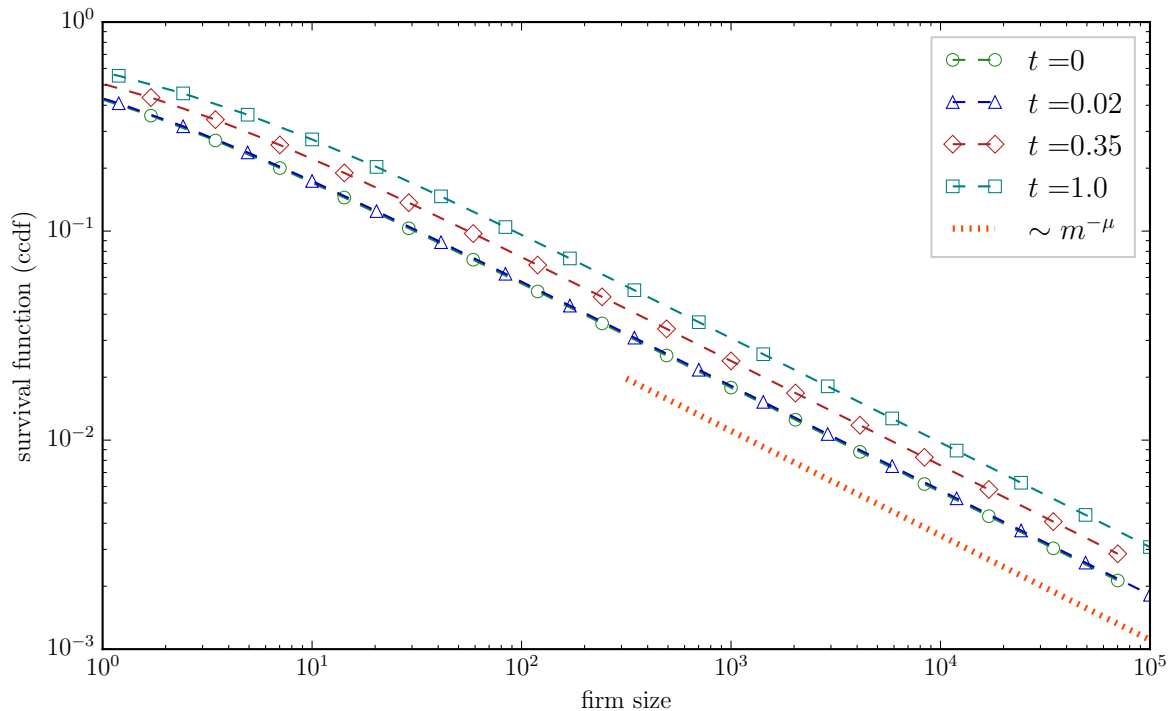


Figure 8.3: The first order solution of the coagulation equation (8.1) with exponential kernel (8.5) and initial distribution (8.7) is given by (8.37) with  $\theta(t)$  determined via (8.44). Here, we show the corresponding survival function (or complementary cumulative distribution function)  $\text{ccdf}(m) \equiv 1 - \int_0^m dm' p(m')$  for  $\mu = 0.5$  and  $\alpha = 0.01$ . The dependence on  $\alpha$  is extremely weak, which is why we do not show different choices. The times  $t$  are chosen in correspondence to the dataset considered by Cefis et al. [183]. All distributions are normalized to one at all times  $t$ .

times  $t < 1$ , only mild deviations from the initial distribution are observed. This is in line with observations 2-4, as we show in the next section.

## 8.4 Connecting the coagulation equation to empirical data

The first order solution to (8.1) was derived in the previous section as (8.37) with  $\theta(t)$  determined by (8.44). So far, we have treated time  $t$  as an abstract parameter. The connection to real time is established by noting that the coagulation kernel  $A(m, m')$  is actually a rate.

Denote the number of firms at time  $t$  by  $N(t)$ . It holds that

$$N(t) = N_0 \int_0^{\infty} dm p(m, t) = N_0 M_0(t) \quad (8.45)$$

with  $N_0 \equiv N(t=0)$ . Then, the total number of firms that have ‘vanished’ through M&A since time  $t=0$  is given by

$$\Delta N(t) = N_0 - N(t) = N_0 (1 - M_0(t)). \quad (8.46)$$

A priori,  $\Delta N < 0$  is also possible, meaning that newly born firms and spin offs outweigh the mergers and increase the total number of firms in an economy. The pure coagulation equation cannot account for such an observation. Here, we do not need to consider this case since 2-4 all report  $\Delta N > 0$ . However, it is, at least in principle, straightforward to extend equation (8.1) to the inclusion of spin-offs and firm births (cf. equation (10.5) in [170]).

Knowing  $M_0(t)$ , equation (8.46) allows us to solve for time  $t$  as a function of observed  $\Delta N$ . In our case, we deduce with (8.43) that

$$t = (1 + \alpha^\mu) \frac{\Delta N(t)}{N_0} \approx \frac{\Delta N(t)}{N_0}. \quad (8.47)$$

In this last approximation, we have used that  $\alpha \ll 1$ . This is justified for the empirical studies below, for which mergers between large companies are reported. We have thus systematically shown that there is negligible dependence on  $\alpha$ , meaning that the kernel  $A$  can be approximated to first order as constant for small times. This result can be made intuitive when noticing that, even for a constant kernel, the far right tail of the distribution is only affected for long times (see also numerical solution in section 8.5). For small times, it is mostly the lower part of the distribution that is affected by coagulation. For small  $m$ ,  $\exp(-\alpha m) \approx 1$  is a good approximation.

Let us now turn to observations 2-4. The most thorough study was conducted by Cefis et al. [183]. They study a comprehensive data set of roughly 60,000 Dutch manufacturing firms including firm entries, exits, spin-offs, mergers and acquisitions. Measuring firm size in number of employees and accounting exactly for all relevant events, Cefis et al. [183] estimate both the firm size distribution in the beginning and in the end of the year 1997. The starting distribution is found to be best fitted by a log-normal. When plotting the starting and final distribution in

one figure, it is seen that the starting and final distribution largely overlap, thus demonstrating that the firm size distribution seems to be unaffected by M&A (cf. Figure 2 there). Since 1997 was a year of high M&A activity of Dutch firms, this observation seems puzzling. While Cefis et al. [183] suggest a number of reasons, this observation finds a natural explanation within our framework. During 1997, 3,899 firms were involved in activities related to M&A and their spin-offs and divestitures. At the end of the year, 2,564 remained, which suggests that M&As, spin-offs and divestitures have decreased the numbers of firms active in the manufacturing sector by  $\Delta N = 1,335$ . With  $N_0 = 57,329$ , we predict from (8.47) a temporal evolution of merely  $t \approx 0.02$ . As is visible in Figure 8.3, the  $p(m, t = 0.02)$  curve is hardly distinguishable from the  $p(m, t = 0)$  curve. The parameter  $t$  is related to real time ( $\approx 365$  days) via a rate, namely the rate of M&A events. The contribution of our analysis is thus to quantify that the M&A events over the full year of 1997 amount in aggregate to a tiny perturbation to the initial firm distribution.

In a next step, in order to emphasize the possible effects of M&As, Cefis et al. [183] restrict their dataset to the firms that are directly involved in a M&A event during the year 1997. The result is a final distribution that is slightly above the starting distribution, showing that the M&As did have an effect on the distribution (cf. Figure 4 there). Again, their result is in good correspondence with our theory. Having  $N_0 = 3,899$  firms in the beginning that will undergo a M&A event and 2,564 of these firms at the end of 1997, we calculate  $\Delta N = 1,335$ . According to (8.47) this corresponds to  $t = 0.35$ . In our model, this results in a distribution that lies also slightly above the initial distribution, as shown in Figure 8.3.

In a very similar fashion, we can explain the observations of Ijiri & Simon [179, 182], at least semi-quantitatively. The major difference between their datasets and the one considered by Cefis et al. [183] is that Ijiri & Simon [179, 182] consider only the tail of the distribution, and do not have access to a complete dataset. In their first study, Ijiri & Simon [182] consider the  $N_0 = 500$  largest industrial firms (in terms of sales volume) and the number of firms that they have bought. They do not find any deviation from Zipf's law through M&A. In our model, this is very natural. Only a total of  $\Delta N = 19$  net mergers are observed, resulting in a small time value of  $t \approx 0.038$ . It is then clear from our calculations that no deviation from Zipf's law is to be expected.

In a follow-up paper, Ijiri & Simon [179] consider a much larger data set and come to the conclusion that M&A does change the firm size distribution by shifting it slightly above from what we would expect without mergers. They do not have a complete dataset, so they provide

an estimate. It is estimated that the 831 largest industrial firms in 1969 would actually be 1,002 companies, if it had not been for a large amount of mergers in the preceding decade. Based on their estimate, we determine  $N_0 = 1,002$  and  $\Delta N = 1,002 - 831 = 171$ , and thus  $t = 0.17$ . The effect that they observe is that the actual curve is shifted slightly to the right of the hypothetical curve without mergers (cf. Figure 1 there). This is again in agreement with what we would expect from our model with  $t = 0.17$ . Furthermore, when plotted in a double logarithmic plot, their observed distributions are concave and thus clearly different from a straight Pareto line, suggesting the lognormal distribution could provide a good fit. That this has no influence on our result becomes clear in the next section, especially from Figs 8.4 (c) and (d).

In conclusion, our model accounts for observations 2-4. Furthermore, we can see that, in order to observe significant deviations from the initial distribution, a large number of mergers and acquisitions (of the order of the number of firms in the dataset) need to be observed, potentially over a long time horizon. This can be traced back to the fact that the heavy tail of the distribution is asymptotically unaffected by the coagulation process with decaying or constant merger kernel. If the firm size distribution was light tailed, deviations in the right end of the distribution would become apparent at an earlier stage, i.e. with fewer M&A observations. See also Figure 8.4 (e) and (f).

## 8.5 Generality of our results

In the previous sections, we have presented a first order analytical solution to the coagulation equation. Thereby, several assumptions and approximations have been made for mathematical convenience. In this section, we show that our results apply more broadly and are not dependent on the assumptions made. Concretely, we use a second-order Runge-Kutta method with variable time step [202] to solve the coagulation equation (8.1) numerically.

Observation 1 states that smaller firms have a higher chance of being absorbed. We have thus assumed an exponentially decaying merger kernel. While this is a natural choice, other shapes are possible. For instance, we could imagine a Pareto kernel  $A(m, m') = 1/m^\beta + 1/(m')^\beta$  for some  $\beta > 0$ . In Figure 8.4, we show the numerical solution to the coagulation equation for both the exponentially decaying kernel ( $A = \exp(-m') + \exp(-m)$ ) and for the constant kernel ( $A = 1$ ). We see that, for  $t < 1$ , the difference between the two kernels is small. Since the Pareto kernel is in between these two extremes of exponential decay and a constant, we conclude that basing our

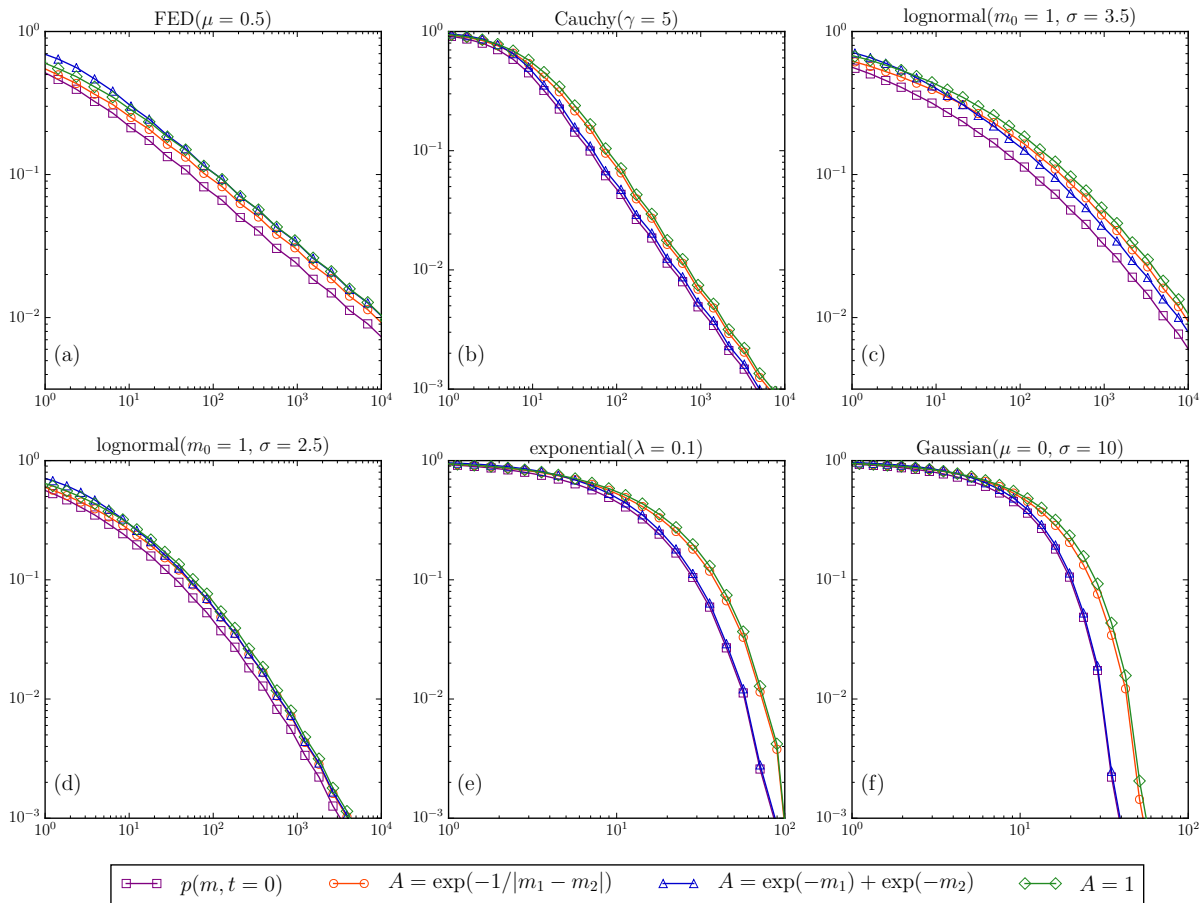


Figure 8.4: Numerical solution, for the complementary cumulative distribution functions (ccdf) of firm sizes, of the coagulation equation (8.1) at time  $t = 1$  for different initial distributions. All distributions are normalized to one at all times  $t$ . We show three different kernels: constant, decaying in company size, and increasing with the difference in company size. Figure (a) corresponds to the initial distribution being a fractional exponential distribution with asymptotic Pareto tail with exponent  $\mu = 0.5$ . We can see that even for the exponentially decaying kernel deviations from the constant kernel solution are small, thus justifying our previous approximations. Similarly, Figure (b) corresponds to the initial distribution being the heavy-tailed Cauchy distribution with shape parameter  $\gamma = 5$  and asymptotic Pareto tail  $\mu = 1$  (Zipf's law). Figure (c) and (d) correspond to the initial distribution being a log-normal law with different values of  $\sigma$  (see definition (8.48)). This confirms that, over a finite size range and for times up to  $t = 1$ , the behavior of the firm distribution as a function of time for the log-normal initial distribution is similar to the case of the power law distributions. Finally, Figure (e) and (f) exemplify that, for light-tailed initial distributions, the dependence on the kernel can appear more pronounced.

calculations on the exponential kernel is justified. In section 8.3, we have also appealed to our intuition that the probability of an M&A increases as a function of the difference in company size. Figure 8.4 addresses this idea by showing a third kernel ( $A = \exp(-1/|m - m'|)$ ). We can see clearly that our results are not affected by this assumption, at least in the limits of heavy tailed initial distributions and small times ( $t < 1$ ).

To simplify the Laplace transform, we have worked with the fractional exponential distribu-

tion with asymptotic power law tail. However, the literature suggests that the lognormal is often also a good approximation of firm size distribution. Empirically, as already mentioned, it can be difficult to distinguish the two. To understand this, note that one can write the log-normal density as

$$p(m, t = 0) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{m} \exp\left(-\frac{\log^2(m/m_0)}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}m_0} (m/m_0)^{-1-\mu(m)} \propto m^{-1-\mu(m)} \quad (8.48)$$

with  $\mu(m) = \frac{1}{2\sigma^2} \log\left(\frac{m}{m_0}\right)$ . The second equality in (8.48) follows because  $e^{(\log m)^2} = m^{\log m}$  (cf. section 4.1.3 in [160]). Since  $\mu(m)$  is a slowly varying function of  $m$ , the log-normal distribution can be mistaken for an apparent power law with an exponent  $\mu$  that is gradually changing. Despite this apparent similarity, the log-normal is very different from a Pareto. According to the Pickands-Balkema-de Haan theorem of extreme value theory [203], the log-normal law belongs to the domain of attraction of the Gumbel distribution for the distribution of maxima, and the Gumbel distribution has an exponential tail. For the log-normal, all moments exist and we would expect that the coagulation equation shows qualitatively different behavior for those two initial distributions (power law versus log-normal). Indeed, Menon and Pego [204, 205] show that (for some specific kernels) the solution of the coagulation equation depends in a subtle way on the number of finite moments of the initial distribution. Nevertheless, as is confirmed in Figure 8.4, these differences are not visible for small times over the considered range of firm sizes. All our theoretical statements seem to apply equally well for other heavy-tailed initial distributions.

## 8.6 Conclusions

We have examined the effect of mergers and acquisitions on time evolution of firm size distributions. Specifically, we have described the effect of M&A in terms of a partial integro-differential equation, also known in the literature as the coagulation equation. This approach allowed us to account for the inverse relation of firm size and the probability of being acquired. While the resulting coagulation equation cannot be solved analytically in closed form, we have derived a first order solution that can explain empirical observations appearing counterintuitive at first sight, namely that mergers and acquisitions develop a significant influence on the firm size distribution only over time scales much longer than a few decades. This explains why M&A has apparently little impact on the firm size distributions in existing data sets. Thus, observations of M&A events over much longer time scales are required in order to see a significant deviation from a

heavy tailed firm size distribution. Precise numerical solutions have confirmed the validity of our approximations. Hence, we conclude that the coagulation equation is an adequate tool to capture the dynamics of merger and acquisitions on the firm size distribution.

We have made two assumptions that could be relaxed. First, we considered the situation where the initial distribution is already heavy tailed. Implicitly, this takes into account the aggregate effect of internal growth, firm births and bankruptcies. These three mechanisms are known to generically create power law distributions of firm sizes [170, 171]. We also note that the effect of mergers and acquisitions was quite negligible at the early stage of a developing economy, as illustrated for the US. Looking at Figure 1 in [178], one sees that M&A have become popular significantly after Gibrat's [158] observation of proportional growth. However, it may be that the asymptotic long time behavior of the firm size distribution is strongly dependent on the initial distribution, a problem that is in full generality still under investigation [196]. It is then interesting to ask whether mergers and acquisitions can significantly impact the firm size landscape of a developing economy.

Second, we have only considered M&As and neglected other firm size dynamics such as internal growth, firm creation, firm exits or sudden bankruptcy. Adding spin-off dynamics to the coagulation equation leads to the more general population balance equation (8.2). In contrast to pure coagulation, here equilibrium states are possible by offsetting the M&As with spin-offs and creation of new firms. An interesting question is then whether such a balance condition is met in real economies, or whether external growth dynamics acts as a driver towards out-of-equilibrium dynamics. Investigation of such questions is part of future research.





## Chapter 9

# Classification of dragon-king phases in preferential attachment and failure models <sup>†</sup>

In classical proportional growth [206] rebranded preferential attachment in the context of networks [207], the probability of an existing node acquiring a new link is only proportional to the number of links it already has (its degree). Consequently, the older the node, the more links it has, which in turn, makes it even more popular (‘the rich get richer’). In highly competitive environments, this assumption appears oversimplified, and one expects competition coming from new and fitter nodes. This has led to the development of fitness-adjusted preferential attachment [208], in which the probability of an existing node acquiring a new link is proportional to the product of its degree and its intrinsic fitness. The more connected and the more fit a node, the more link it gets (‘the fit get richer’). This type of attachment rule has been justified as the result of an evolutionary mechanism that minimizes the maximum exposure to nodes of low fitness [209]. It has been shown [210–212] that, for some distributions of fitnesses in the limit of infinite systems, one or several nodes may dominate the system in the sense that they capture a macroscopic fraction of all links. In a socio-economic context, we call such nodes dragon-kings (DKs) [213,214], since they are endogenously generated statistical outliers beyond the power law regime that prevails in absence of any such DKs.

In this chapter, we add an additional ingredient to the picture: preferential removal of nodes. At any instance of time, a node may be removed from the system with some probability, e.g. due

---

<sup>†</sup>This chapter is based on reference [16].

to failure, old age, attacks, etc. We consider this probability as proportional to the node fitness, raised to some exponent  $\omega$  that may be positive or negative. This is a generalization of previous works, which have considered either uniform deletion rates ( $\omega = 0$ ) [215], or preferential deletion that is inversely proportional to the node degree, but without node fitnesses [216]. Additionally, we provide an exact condition for the emergence of DK nodes in this system, thereby extending results known for pure preferential attachment [210–212]. Since the phase-transition is dependent on the distribution of fitnesses in the network, we then ‘parametrize’ the space of fitnesses through the beta-distribution, allowing us to draw phase-diagrams, and characterizing the fraction of DK nodes in the system.

The network topology of the generated graph can be described in terms of the survival probability and size distribution of a multitype branching random walk [217]. Admittedly a bit far fetched, this analogy enables us to embed the current chapter under the title of a branching (and hence ‘constrained’) stochastic process. The applicability of preferential attachment models to socio-economic systems have been shown repeatedly, see for instance [209, 218].

## 9.1 Theoretical framework of network growth with fitness-adjusted preferential attachment and deletion

Consider growing an undirected network by attaching one node per unit time to an already existing network. At time  $t = 0$ , we start with an initial network consisting of  $\ell$  nodes and  $e_0$  edges. At each time step  $t = 1, 2, 3, \dots$ , a new node is added to the network by attaching it to  $m$  randomly chosen vertices. At time  $t$ , there are then  $\ell + t$  nodes  $n_1, n_2, \dots, n_{\ell+t}$  and  $mt + e_0$  edges in the network. In the limit of large  $t$ , we can safely approximate  $\ell + t \approx t$  and  $mt + e_0 \approx mt$ . This approximation is used in the remainder of this chapter. We assume that each node  $n_i$  has an intrinsic fitness  $\eta_i \in (0, 1)$  sampled from some fitnesses distribution  $\rho$ . Denoting by  $k_i$  the number of edges connected to node  $n_i$  (i.e. its degree), we assume that an edge of a newly introduced node attaches to  $n_i$  with probability proportional to both the product of its degree and its fitness. The fitter and the more connected a node, the more attractive it is.

Nodes may also be removed from the system (failure, death, etc.). At each time step, with probability  $c$ , a randomly selected node is removed from the network, such that, at time  $t$ , there are on average  $N(t) = (1-c)t$  nodes in the system. It is reasonable to assume that the probability of failure depends on the fitness of the node. Specifically, we assume that the probability

that node  $n_i$  fails is proportional to  $\eta_i^\omega$  for some exponent  $\omega$ . If  $\omega > 0$ , then the fit nodes are more likely to fail. This may be for instance because fit nodes are subject to targeted attacks. On the other hand, if  $\omega < 0$ , then weak nodes are likely to fail. This is reasonable for instance if we interpret  $\eta$  itself as a measure of robustness with respect to failure.<sup>1</sup> The special case  $\omega = 0$  of uniform failure rates has been considered in ref [215].

Assuming that the  $i$ -th node is still alive at time  $t$ , the rate at which the degree of node  $n_i$  increases on average is

$$\frac{\partial k_i}{\partial t} = m \frac{\eta_i k_i}{S(t)} - c k_i \sum_{\text{nbrs } j} \frac{\eta_j^\omega}{N \langle \eta^\omega \rangle}, \quad (9.1)$$

where

$$S(t) = \int_1^t dj D(j, t) k_j(t) \eta_j \quad (9.2)$$

is a normalization factor and  $D(j, t)$  is the probability that the  $j$ -th node is still present in the system at time  $t$ . The first term on the right hand side of equation (9.1) is a direct consequence of the attachment rule described above, and  $m$  accounts for the  $m$  edges of the new node. The second factor is a product of two independent probabilities. The probability  $c$  that a randomly chosen node is deleted, and the probability that node  $n_i$  is connected to the randomly failed node. The sum in (9.1) extends over all neighbors of node  $n_i$  and we have normalized according to  $\sum_\ell \eta_\ell^\omega = N \frac{1}{N} \sum_\ell \eta_\ell^\omega \approx N \langle \eta^\omega \rangle$ .

The sum depends on the specific environment of node  $n_i$ . We consider a mean-field version of this equation by replacing the configuration specific value of  $\eta_j$  entering in  $\eta_j^\omega$  in the second term of the r.h.s. of equation (9.1) by the average node fitness  $\langle \eta \rangle$ , where the average is over the fitness distribution  $\rho$ . The mean-field version of (9.1) then becomes

$$\frac{\partial k_i}{\partial t} = m \frac{\eta_i k_i}{S(t)} = \underbrace{\frac{c}{1-c} \frac{\langle \eta \rangle^\omega}{\langle \eta^\omega \rangle}}_{\equiv \Omega > 0} \frac{k_i}{t}, \quad (9.3)$$

where we have used that

$$N(t) \approx (1-c)t. \quad (9.4)$$

---

<sup>1</sup> Alternatively, we could also assume that the failure rate is proportional to  $(k_j \eta_j)^\omega$ , at the cost of making the following calculations much harder to track analytically. Nevertheless, since high fitness nodes, by construction, tend to be well connected, we may assume that the two models capture qualitatively similar behavior.

We make a multi-scaling ansatz for the solution of (9.3), imposing that

$$k_i = m \left( \frac{t}{i} \right)^{\gamma(\eta_i)}. \quad (9.5)$$

The scaling exponent  $\gamma$  depends on the fitness  $\eta_i$  of the node. This is motivated for instance through the well known-result [206] that in classical preferential attachment (no node removal, uniform fitnesses) the degree of node  $n_i$  scales like  $k_i \sim t^{1/2}$ .

Next, we derive an explicit expression for  $D(i, t)$ , the probability of node  $n_i$  still being alive at time  $t$ . Clearly, it holds iteratively that

$$D(i, t + 1) = D(i, t) \cdot \left[ 1 - c \frac{\eta_i^\omega}{\sum_j \eta_j^\omega} \right], \quad (9.6)$$

which, in the continuous limit, yields

$$\frac{\partial D(i, t)}{\partial t} = - \underbrace{\frac{c}{1-c} \frac{1}{\langle \eta^\omega \rangle}}_{\equiv \kappa > 0} \eta_i^\omega \frac{D(i, t)}{t}, \quad (9.7)$$

where we have again used (9.4). Together with the initial condition  $D(t, t) = 1$  (the node that is just introduced cannot be deleted in the same time-step), we find

$$D(i, t) = \left( \frac{t}{i} \right)^{-\kappa \eta_i^\omega}. \quad (9.8)$$

Next, we want to determine  $S$ . Plugging (9.8) into (9.2) yields

$$S(t) = m \int_1^t di \left( \frac{t}{i} \right)^{\gamma(\eta_i) - \kappa \eta_i^\omega} \eta_i. \quad (9.9)$$

The integration in (9.9) involves the  $\eta_i$ 's, which are stochastically sampled from the distribution  $\rho(\eta)$ . We consider large times  $t$  for which the integral in (9.9) is dominated by large  $i$ 's, as verified consistently below. Considering therefore the large values of  $i$  in the integral, let us take the integration over some interval  $[i, i + di]$  such that  $di/i$  is small but  $di$  is large. The integral of  $i$  in this interval  $[i, i + di]$  is then equivalent to sampling the random variables  $\eta_i$ 's according to their distribution  $\rho(\eta)$ . Given that  $\frac{t}{i}$  can be considered approximately constant in the interval

$[i, i + di]$ , expression (9.9) becomes

$$\langle S(t) \rangle = m \int_0^1 d\eta \rho(\eta) \eta \int_1^t di \left( \frac{t}{i} \right)^{\overbrace{\gamma(\eta) - \kappa\eta^\omega}^{\tilde{\gamma}}} \quad (9.10a)$$

$$= m \int_0^1 d\eta \rho(\eta) \eta \frac{t - t^{\tilde{\gamma}}}{1 - \tilde{\gamma}}. \quad (9.10b)$$

Now, we argue that  $\tilde{\gamma} = \gamma(\eta) - \kappa\eta^\omega < 1$ . To understand why this must hold, it is easier to think first of a network where there is no deletion, i.e.  $c = 0$  and  $\tilde{\gamma} = \gamma$ . Since every node, by construction, will have a fitness  $\eta > 0$  and at least  $m$  links, it will always have a finite probability of attracting more links. Hence, its node degree will always be increasing, implying that  $\gamma > 0$ . On the other hand, in the most extreme case, node  $n_i$  is so strong that it captures a new link every time a new node is added to the system. In that case, the node degree grows linearly,  $\gamma = 1$ . Since, by construction, we do not have such a scenario under ‘normal circumstances’, we conclude  $\gamma < 1$ . This constraint breaks down in the case of a phase transition with the emergence of DK nodes, where  $\tilde{\gamma} = 1$  for the fittest nodes. If we allow for node deletion, we must replace the above reasoning for  $k_i(t)$  by the expected degree of node  $n_i$  at time  $t$  [219]. Since, by construction, a node that was removed from the network has zero edges, this quantity is given by  $\mathbb{E}[k_i] = k_i \cdot D(i, t) \propto t^{\gamma - \kappa\eta^\omega} = t^{\tilde{\gamma}}$ . With the same reasoning as for the case without node deletion, we conclude that  $\tilde{\gamma}$  must be below 1. Note that, somewhat counter-intuitively, the requirement  $\tilde{\gamma} < 1$  still allows for  $\gamma > 1$ . In other words, the individual node may grow super-linearly, but only for a short time, as its lifetime is also limited [215, 219].

Thus, in the limit  $t \rightarrow \infty$ , the term  $t^{\tilde{\gamma}}$  is negligible with respect the term linear in  $t$ , and we end up with

$$\langle S(t) \rangle \approx Amt, \quad (9.11)$$

where

$$A = \int_0^1 d\eta \frac{\eta \rho(\eta)}{1 - \tilde{\gamma}}. \quad (9.12)$$

Plugging this asymptotic limit into (9.3) yields

$$\frac{\partial k_i(\eta)}{\partial t} = \left( \frac{\eta}{A} - \Omega \right) \frac{k_i(\eta)}{t}. \quad (9.13)$$

The solution of this equation is exactly of the imposed form (9.5), with

$$\gamma(\eta) = \frac{\eta}{A} - \Omega. \quad (9.14)$$

Finally, the constant  $A$  remains to be determined by self-consistently plugging (9.14) into (9.12), leading to the condition

$$1 = \int_0^1 d\eta \frac{\rho(\eta)}{A \left( \frac{1+\Omega}{\eta} + \frac{\kappa}{\eta^{1-\omega}} \right) - 1}. \quad (9.15)$$

This ends the general discussion, and we now have to distinguish between two different cases: The case where equation (9.15) has a solution and the case where it does not.

## 9.2 Generic case: The fit get richer

In case there is an  $A$  solving equation (9.15), the cumulative degree distribution  $P_\eta$  for fixed  $\eta$  is derived through

$$P_\eta [k_i(\eta) < k] = P_\eta \left[ m \left( \frac{t}{i} \right)^\gamma < k \right] \quad (9.16a)$$

$$= P_\eta \left[ i > \left( \frac{m}{k} \right)^{1/\gamma} t \right] \quad (9.16b)$$

$$= 1 - P_\eta \left[ i \leq \left( \frac{m}{k} \right)^{1/\gamma} t \right] \quad (9.16c)$$

$$= 1 - \frac{1}{1-c} \left( \frac{m}{k} \right)^{1/\gamma}. \quad (9.16d)$$

In the last step, we have used that the probability of choosing any node  $n_i$  at random is uniform, and there are  $(1-c)t$  nodes in the network according to expression (9.4). The pdf is then obtained by taking the derivative, giving  $p_\eta \propto k^{-\left(1+\frac{1}{(1-c)\gamma(\eta)}\right)}$ . The overall pdf is then just

$$p(k) \propto \int_0^1 d\eta \rho(\eta) k^{-\left(1+\frac{1}{(1-c)\gamma(\eta)}\right)}. \quad (9.17)$$

Different fitness distributions give rise to different power law exponents through  $\gamma$ , that are asymptotically dominated by  $\max_\eta \gamma(\eta)$ , as is made intuitive through a saddle-point approximation, see for instance [215, 219]. We call this case the ‘scale-free’ phase, in contrast to the case discussed next, in which some of the fittest nodes completely dominate the system.

### 9.3 Special case: Dragon-king nodes

It may happen that equation (9.15) has no solution for  $A$ . To gain an understanding of what this means, note that  $S(t)$  has the interpretation of the (fitness adjusted) number of nodes in the system at time  $t$ , and should hence scale linearly with  $t$ , as illustrated by expression (9.11). Specifically, if there is no removal of nodes ( $c = 0$ ) and all fitnesses are equal,  $S(t)$  reduces to the total number of outgoing links in the system,  $S(t) = 2mt$  and  $A = 2$ . If nodes have varying fitnesses and may be removed,  $A$  has to be adjusted. But if (9.15) cannot be satisfied, this means no such  $A$  exists. Going again through the above derivation, one notices that the only assumption that could have been violated is  $\tilde{\gamma} < 1$ . If  $\tilde{\gamma} = 1$  for some nodes, they attract links proportional to the number of nodes added to the system. In a system of infinite size, they still connect a macroscopic fraction of all nodes in the system. They are dragon-king nodes [213,214].

This intuitive reasoning has been confirmed more formally for  $c = 0$ , using a mapping of the network to a Bose-Einstein condensate [210]. In that mapping, the fitness  $\eta$  is mapped to an energy level  $\epsilon = -\log(\eta)/\beta$  and  $\beta$  is the inverse temperature. The associated distribution  $g(\epsilon) = \beta e^{-\beta\epsilon} \rho(\exp(-\beta\epsilon))$  can then be interpreted as the normalized degeneracy of the energy level  $\epsilon$ . While elegant, it is not always straightforward to see how this mapping can be extended to different conditions, or arbitrary fitness distributions. For instance, in case the fitness distribution has an integrable singularity at  $\eta \uparrow 1$  (c.f. the  $\alpha = \beta = 1/2$  line in Figure 9.1), the degeneracy distribution  $g$  has a divergence at  $\epsilon = 0$ , i.e. an infinitely degenerate ground-state, with no clear interpretation. Other derivations relying on master- or rate equations have been developed [211,212], but to the best of our knowledge not in combination with preferential deletion.

Here, we therefore derive a systematic classification of the existence of DK nodes that does not rely on a mapping to Bose-Einstein statistics. Specifically, we derive an exact condition for equation (9.15) to have no solution. To this end, we define

$$I(A) \equiv \int_0^1 d\eta \frac{\rho(\eta)}{A \left( \frac{1+\Omega}{\eta} + \frac{\kappa}{\eta^{1-\omega}} \right) - 1}, \quad (9.18)$$

and ask if there is an  $A$  that satisfies  $I(A) = 1$ . To answer this question, we look for a lower boundary  $\underline{A}$  for which we know that definitely no  $A < \underline{A}$  with  $I(A) = 1$  exists. Our claim is then that DK nodes are present in the system if and only if  $I^* \equiv I(\underline{A}) < 1$ . To see this, note that  $I(A)$  is continuous and monotonically decreasing from  $I^*$  to zero as a function of  $A$ . Assuming

that  $I^* \geq 1$  but finite, we can then continuously increase  $A$  from  $A = \underline{A}$  for which  $I(A = \underline{A}) > 1$  to a larger value until  $I(A) = 1$  is satisfied. Conversely, if  $I^* < 1$ , we cannot find any  $A$  with  $I(A) = 1$  since the function  $I(A)$  is decreasing with  $A$ . A case that needs special consideration is when  $I^*$  diverges. However, in that case, it is easy to see that we can always find some  $A_1 > \underline{A}$  such that  $I(A_1) > 1$  and is finite. Indeed, by continuity, we have  $\lim_{A_1 \downarrow \underline{A}} I(A_1) = +\infty$ . On the other hand, we can always choose an  $A_2$  large enough such that the denominator in (9.18) is larger than 1 for all  $\eta \in (0, 1)$ . Then,  $I(A_2) < \int d\eta \rho = 1$  by normalization. Therefore, an  $A \in (A_1, A_2)$  exists with  $I(A) = 1$ . This concludes the proof of our claim.

We have thus reduced the question of the existence of a solution for equation (9.15) to the determination of the lower boundary  $\underline{A}$ . Note that, if the denominator of the integrand of (9.18) has a zero value on its integration domain,  $I(A)$  is likely to diverge for general  $\rho$ 's, and hence cannot satisfy  $I(A) = 1$ . Denote by  $\eta_{\min} \in (0, 1]$  the  $\eta$  that minimizes the expression  $(1 + \Omega)/\eta + \kappa/\eta^{1-\omega}$ . Then, the lower bound  $\underline{A}$  is determined as

$$\underline{A} = \left( \frac{1 + \Omega}{\eta_{\min}} + \frac{\kappa}{\eta_{\min}^{1-\omega}} \right)^{-1} = \frac{\eta_{\min}}{1 + \Omega + \kappa\eta_{\min}^{\omega}}, \quad (9.19)$$

with

$$\eta_{\min} = \begin{cases} \left( \frac{1+\Omega}{\kappa(\omega-1)} \right)^{1/\omega} & \text{if } \langle \eta \rangle^{\omega} + \frac{1-c}{c} \langle \eta^{\omega} \rangle < \omega - 1 \\ 1 & \text{else.} \end{cases} \quad (9.20)$$

To wrap up, we put  $\underline{A}$  given by (9.19) into (9.18) and conclude that there are DK nodes (no solution of equation (9.15) for  $A$  can be found) if and only if

$$I^* = \int_0^1 d\eta \frac{\rho(\eta)}{\frac{\eta_{\min}}{1+\Omega+\kappa\eta_{\min}^{\omega}} \left( \frac{1+\Omega}{\eta} + \frac{\kappa}{\eta^{1-\omega}} \right) - 1} < 1, \quad (9.21)$$

with  $\eta_{\min}$  given by (9.20).

For  $\omega = 0$ , condition (9.21) simplifies to  $\int_0^1 d\eta \rho \left( \frac{1}{\eta} - 1 \right)^{-1} < 1$ , which is exactly the condition derived in ref. [210] through the mapping to the Bose-Einstein gas. Note that this holds even in case of node deletion  $c > 0$ . The interpretation is that when nodes are removed uniformly at random, this has no influence on the phase-transition.



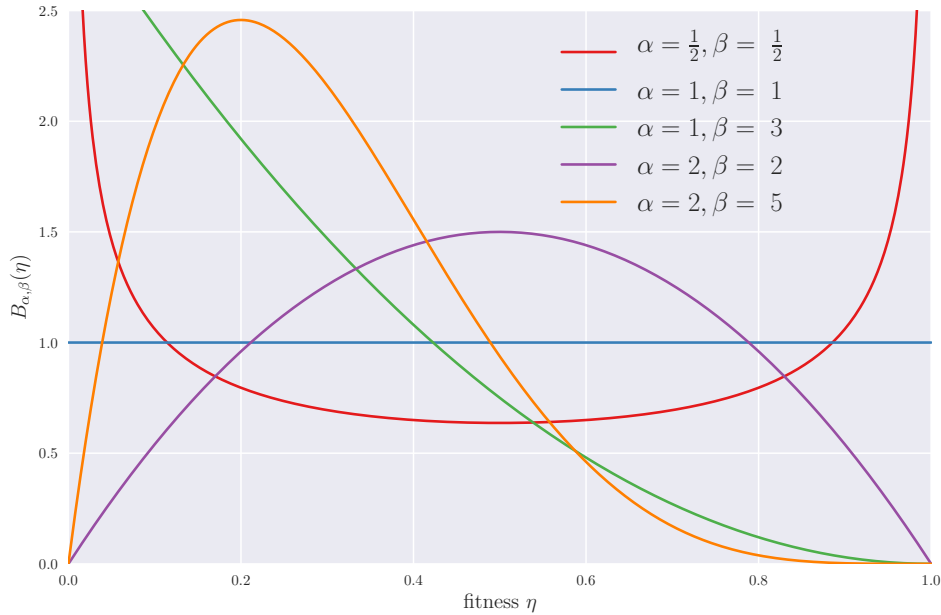


Figure 9.1: Beta distribution (9.22) for different set of parameters  $\alpha$  and  $\beta$ . We assume that this parametric class of distribution approximates the space of fitness distributions sufficiently well for all practical cases. DK nodes emerge when the bulk of the probability mass is concentrated on nodes of low fitness, and there are just a few fit nodes, that then end up dominating the system (e.g.  $\alpha = 2, \beta = 5$ ).

## 9.4 Parametric classification of the DK regime

In the previous section, we have seen that, for certain fitness distributions and deletion rates, DK nodes emerge and completely dominate the system. But can we somehow classify such distributions? Clearly, there is an uncountably infinite number of fitness distributions  $\rho(\eta)$  on the interval  $[0, 1]$ . While it is impossible to characterize such distributions systematically, we propose to work with a rich family, the beta-distribution, with its density given by

$$\beta(\eta; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} \eta^{\alpha-1} (1 - \eta)^{\beta-1} \quad (9.22)$$

and  $B$  the normalizing beta-function. Given the various shapes that are obtained when tuning its parameters (Figure 9.1), we claim that, to first approximation, it captures all practically relevant classes of fitness constellations. In a sense, we thus ‘parametrize’ the space of fitness distributions. This has the advantage that the DK node condition (9.21) is now a function of four parameters  $\{\alpha, \beta, \omega, c\}$ . The value of  $I^*$  is thus a function  $I^*(\alpha, \beta, \omega, c)$  of these four parameters, which determine the phase transition from scaling behavior to DK occurrence.

In general, the integral in (9.21) with  $\rho$  given by the beta-density (9.22) has to be solved

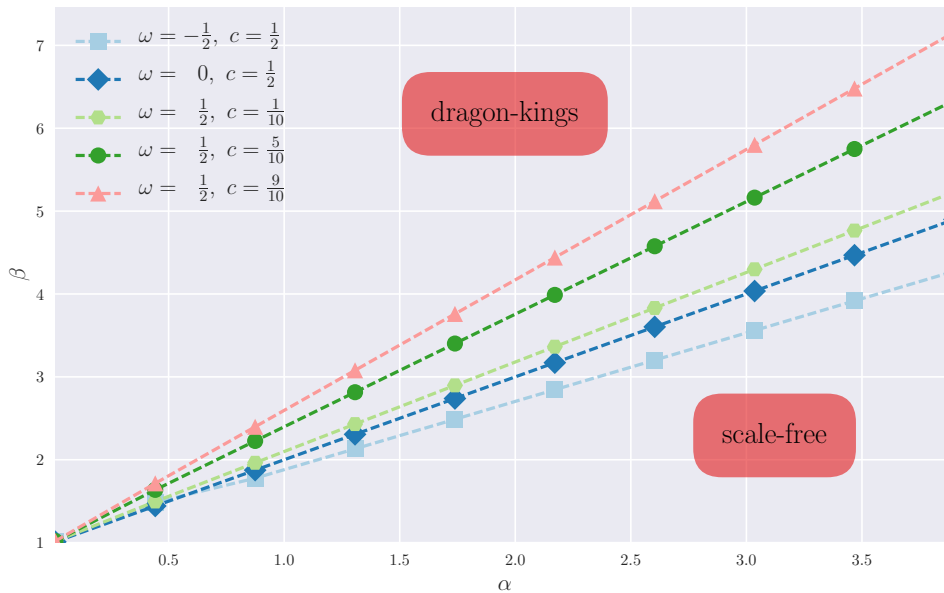


Figure 9.2: Phase-diagram of the scale-free vs. dragon-king phases in the  $(\alpha, \beta)$ -parameter plane of the beta-distribution, for different preferential deletion parameters  $\omega$  and  $c$ . The scale-free regime is the one below the separation line(s), corresponding to fitness distributions that tend to be either balanced or overemphasize the fit nodes. In contrast, if the distribution is such that most probability mass is concentrated on nodes with low fitness, dragon-king nodes can emerge. This effect is enhanced the smaller  $\omega$  and the larger  $c$ , i.e. the more nodes of low fitness are removed. In contrast, if  $\omega$  is large, the scale-free regime is widened, as the targeted attacks to fit nodes tend to remove dragon-kings. The larger  $c$ , the stronger this effect.

numerically. However, the simplified case  $\omega = 0$  of uniform node deletion allows for an explicit result:

$$I^* = \begin{cases} \frac{\alpha}{\beta-1}, & \text{if } \beta > 1 \\ +\infty, & \text{if } \beta \leq 1. \end{cases} \quad (9.23)$$

We thus conclude that, for  $\omega = 0$  with  $\alpha < \beta - 1$  and  $\beta > 1$ , it holds that  $I^* < 1$  and hence there are DK present. The separation between the scale-free and the DK-regime is along the line  $\beta = \alpha + 1$ . Figure 9.2 depicts the regime plot for general parameter configurations. Comparing with Figure 9.1, we see that the DK regime corresponds to fitness landscapes in which the bulk of the nodes have low fitnesses, such that the few fit nodes end up dominating the system. In line with our intuition, the larger  $\omega$ , the more numerous are the fitness distributions that fall into the scale-free regime. This effect is enhanced, the larger the value of the deletion rate  $c$ .

In conclusion, we have extended the calculations of fitness-adjusted preferential attachment by also considering fitness dependent preferential deletion. We have shown under what circumstances the system undergoes a phase transition where some nodes capture a finite fraction of all links in the infinite network. By parametrizing the space of fitness distributions through the

beta-density, we could furthermore derive explicit analytical results that characterize the regime change as a function of four parameters.



# Chapter 10

## Conclusions

We have considered three different types of constrained stochastic processes, applied to different financial and economic data. In the first part, we have focused on currency target zones. In an initial, slightly naive attempt of chapter 2, we have described such target zones from the perspective of statistical physics, as the combination of a free random walk and a reflective barrier, giving rise to an effective repulsive force. Non-parametric estimates of the EUR/CHF target zone were in contradiction with that theory, showing that the drift term is essentially vanishing, whereas the volatility is decaying like a square-root as a function of the distance to the barrier. This made us realize that if such a repulsive force was present, significant arbitrage opportunities would prevail that are completely unrealistic from a financial point of view. All non-trivial behavior is thus shifted into the dynamics of the volatility. We have then turned our attention to a previously established result stating that the exchange rate price can be considered as the midpoint of an effective ‘order book fluid’. Building on that analogy, we have finally derived that square-root behavior of the volatility from first order principles of hydrodynamics, showing that the order-book analogy extends to the case of a rigid constraint.

In a second approach of chapter 3, we have turned our attention to economic theories, in particular to the famous target zone model by Krugman. In contrast to the naive approach from the previous chapter, here the no-arbitrage criterion is fundamentally incorporated into the theory as a parameter, and we have applied statistical tests to check that indeed arbitrage was unlikely for the recent euro/Swiss-franc target zone. Using first order approximations, we could derive exactly the square-root shaped volatility that we have observed empirically, and likelihood ratio tests would prove the fit significant. In conclusion, we have shown that the classic target zone model by Krugman, famous for its simplicity, provides an accurate description of the

euro/Swiss-franc target zone, as it represents a case of an exceptionally strictly enforced one.

Instead of the phenomenological modeling approach by Krugman, target zones can alternatively be described in terms of put and call options. So far, only single-sided target zones have been considered from that perspective. We have thus extended the mathematical formulation to general two-sided target zones in chapter 4, showing that they are accurately represented in terms of an infinite sequence of nested local volatility put and call options. Using sophisticated numerical algorithms, we could show that the sequence is converging, but iterations up to fifty orders or higher are required for satisfactory numerical precision, as demonstrated on the US dollar/Hong-Kong dollar exchange rate. We could then use this representation to estimate the unobservable underlying fundamental value from the observable exchange rate, using maximum likelihood techniques. By interpreting the option maturity parameter as a proxy for the perceived market credibility and estimating it from the data, we have furthermore provided a dynamic estimate for the future duration of the target zone policy. The advantage of this option model is that the effectiveness of different target zone policies can conveniently be assessed, by simply replacing the American options considered in this thesis with more exotic ones. As part of future research, we can consider for instance forward-looking target zones, market inefficiencies, or soft target zone boundaries by pricing appropriate options.

Chapter 5 has finally closed the gap between the phenomenological modeling approach from chapter 3 and the option approach from chapter 4. By non-linearly generalizing the model of Krugman, we could map it in a one to one manner to the options model, at least for the case of perpetual options and single-sided target zones. The mapping allowed us to express the empirically difficult to measure semi-elasticity parameter of the phenomenological literature in terms of economically well-defined quantities, thereby resolving a long standing problem regarding the numerical estimation of this parameter. The extension to more general set-ups is left for future research.

In a second part of the thesis, we have shifted our attention from intraday forex data to real, US GDP growth rate data over the past 200 years. Chapter 6 can be seen as a precursor to chapter 7, in which we have applied sequentially a wavelet transform and a subsequent statistical test in order to assess the bimodality of the GDP growth rate distribution. We rationalized our findings in terms of the social-bubble hypothesis, stating that the economy is fundamentally out of equilibrium, oscillating between periods of high growth (booms) and low or even negative growth (recessions). As a side result, we have contrasted this approach to previous parametric

approaches, that would pre-condition their results based on economic principles that could not be confirmed in our analysis. Specifically, it was claimed that business cycles are naturally restricted to a certain range of periodicities, with seasonalities exceeding that range characterized as noise on the one side and long term trends on the other. Our non-parametric continuous approach did not support this claim. Instead, we have found a continuous hierarchy of seasonalities at all scales, that smoothly renormalize into the observed long-term growth of roughly 2%.

The findings from GDP data were complemented in chapter 7 by showing that the US GDP growth rate distribution is not only bimodal, but also possesses symmetric, heavy tails. We have then presented a simple model that non-trivially combines these two observations. Within that model, the GDP growth rates are the result of a Lévy flight embedded in a steep (more than quadratic) potential well. The associated non-linear restoration force can be seen as a renormalized representation of the dampening effect that the overall economic interaction network exhibits on the idiosyncratic growth rate shocks to individual firms or sectors. Via bootstrapping, we have shown that with merely 200 datapoints, all five model parameters can be estimated fairly accurately, allowing us to fit the model to the 200 annual GDP growth rate datapoints. We found clear support for the model, with potential walls that are just slightly steeper than quadratic. As part of future research, the model can be furnished with additional ingredients, in particular correlated noise, to account even better for the momentum associated with the social-bubble hypothesis mentioned above.

In chapter 8, we have turned our attention to firm size distributions, which are well-known to be heavy-tailed. Previous work has examined empirically the effect of merger and acquisitions on the tail exponent of that distribution, finding that there is no significant effect. To explain this somewhat surprising result, we have set up an integro-differential equation that captures the firm interaction through merger and acquisitions explicitly. Solving that equation both analytically to first order and numerically for various initial conditions, we could show that the number of merger and acquisitions considered in the empirical studies was not enough to observe a significant effect. We have attributed this to the heavy tail of the distribution. If it was lighter, a clear impact from merger and acquisitions would be immediately apparent. As part of future work, we extend our model to also take into consideration firm birth, bankruptcy, spin-offs, internal growth and other ingredients that may affect the firm size distribution.

The final chapter 9 has been devoted to the presentation of a new type of preferential attachment model. As opposed to classical preferential attachment, here the probability of an existing

node acquiring a new link is only proportional to the number of links it already has multiplied by its intrinsic fitness. For some distributions of fitnesses, in the limit of infinite systems, one or several nodes may dominate the system in the sense that they capture a non-zero fraction of all links (dragon-kings). By ‘parametrizing’ the space of fitness distributions through the Beta-distribution, we were able to classify exactly what type of fitness landscapes give rise to such kind of outliers that are destabilizing the system. To further account for the finite lifetime of any node, we add additionally a preferential deletion mechanism that, depending on the parameters, may prefer especially fit or especially weak nodes. As part of future work, we will fit the model to actual data, determine the position of the observed fitness landscape within the phase diagram and apply slight perturbations if necessary, in order to prevent the formation of dragon-kings.



# Bibliography

- [1] W. Paul and J. Baschnagel. *Stochastic processes - From Physics to Finance*. Springer, 2013.
- [2] D. Sornette. Physics and financial economics (1776-2014): puzzles, Ising and agent-based models. *Reports on Progress in Physics. Physical Society (Great Britain)*, 77(6):062001, 2014.
- [3] D. Sornette. Brownian representation of fractal quantum paths. *European Journal of Physics*, 11(6):334, 1990.
- [4] L. Bachelier. *Théorie de la spéculation*. Gauthier-Villars, 1900.
- [5] P.A. Samuelson. Proof that Properly Anticipated Prices Fluctuate Randomly. *Industrial Management Review*, 6(2):41–49, 1965.
- [6] Y. Yura, H. Takayasu, D. Sornette, and M. Takayasu. Financial brownian particle in the layered order-book fluid and fluctuation-dissipation relations. *Physical Review Letters*, 112(9):1–5, 2014.
- [7] Y. Yura, H. Takayasu, D. Sornette, and M. Takayasu. Financial Knudsen number: Break-down of continuous price dynamics and asymmetric buy-and-sell structures confirmed by high-precision order-book information. *Physical Review E*, 92(4):1–12, 2015.
- [8] J.-P. Bouchaud, Y. Gefen, M. Potters, and M. Wyart. Fluctuations and response in financial markets: The subtle nature of random price changes. *Quantitative Finance*, 4(2):176–190, 2004.
- [9] S. C. Lera and D. Sornette. Currency target-zone modeling: An interplay between physics and economics. *Physical Review E*, 92(6):062828, 2015.
- [10] S. C. Lera and D. Sornette. Quantitative modelling of the EUR/CHF exchange rate during the target zone regime of September 2011 to January 2015. *Journal of International Money and Finance*, 63:28–47, 2016.
- [11] S. C. Lera, M. Leiss, and D. Sornette. Currency Target Zones as Mirrored Options. *SSRN*, (Working Paper No. 170), 2017.
- [12] S. C. Lera and D. Sornette. An Explicit Mapping of Currency Target Zone Models to Option Prices . *SSRN*, (Working Paper Series), 2017.

- [13] S. C. Lera and D. Sornette. Evidence of a Bimodal US GDP Growth Rate Distribution : a Wavelet Approach. *Quantitative Finance and Economics*, 1(1):26–43, 2017.
- [14] S. C. Lera and D. Sornette. Gross domestic product growth rates as confined Lévy flights: Towards a unifying theory of economic growth rate fluctuations. *Physical Review E*, 97(1):12150, 2018.
- [15] S. C. Lera and D. Sornette. Quantification of the evolution of firm size distributions due to mergers and acquisitions. *PLoS ONE*, 12(8):e0183627., 2017.
- [16] S. C. Lera and D. Sornette. Classification of dragon-king phases in preferential attachment and failure models. *arXiv preprint arXiv:1711.09890*, 2017.
- [17] A. Einstein. Über die von der Molekularkinetischen Theorie der Wärme geforderte Bewegung von in Ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 322:549–560, 1905.
- [18] J. A. Frankel. No Single Currency Regime is Right for all Countries or at all Times. *NBER Working Paper Series*, (Working Paper 7338), 1999.
- [19] E. F. Fama. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2):383–417, 1970.
- [20] E. F. Fama. Efficient capital markets: II. *The journal of finance*, 46(5):1575–1617, 1991.
- [21] E. F. Fama. Two pillars of asset pricing. *The American Economic Review*, 104(6):1467–1485, 2014.
- [22] J.-P. Bouchaud and Cont R. A Langevin Approach to Stock Market Fluctuations and Crashes. *The European Physical Journal B*, 6(4):543–550, December 1998.
- [23] J. D. Farmer. Market Force, Ecology and Evolution. *Industrial and Corporate Change*, 11(5):895–953, 2002.
- [24] K. Ide and D. Sornette. Oscillatory Finite-Time Singularities in Finance, Population and Rupture. *Physica A*, 307(1-2):63–106, 2002.
- [25] S. Chandrasekhar. Stochastic problems in physics and astronomy. *Reviews of modern physics*, 15(1):1–89, 1943.
- [26] S. Redner. *A Guide to First-Passage Processes*. Cambridge University Press, 2001.
- [27] M.E. Fisher. Walks, walls, wetting, and melting. *Journal of Statistical Physics*, 34(5):667–729, 1984.
- [28] D. Sornette and N. Ostrowsky. Lamellar phases: Effect of fluctuations (theory). In *Micelles, Membranes, Microemulsions and Monolayers*, pages 251–302. Springer New York, 1994.
- [29] C. C. Chow and M. A. Buice. Path integral methods for stochastic differential equations. *The Journal of Mathematical Neuroscience*, 5(1):8, 2015.

- [30] S. Lera. *Constrained Random Walk Models*. Master thesis, ETH Zurich, 2015.
- [31] H. Risken. *The Fokker-Planck equation*. Lecture Notes in Mathematics. Springer Berlin-Heidelberg, 1996.
- [32] R. Friedrich, S. Siegert, J. Peinke, St Lück, M. Siefert, M. Lindemann, J. Raethjen, G. Deuschl, and G. Pfister. Extracting model equations from experimental data. *Physics Letters, Section A: General, Atomic and Solid State Physics*, 271(3):217–222, 2000.
- [33] P. Sura and J. Barsugli. A note on estimating drift and diffusion parameters from timeseries. *Physics Letters A*, 305:304–411, 2002.
- [34] H. Brenner. The slow motion of a sphere through a viscous fluid towards a plane surface. *Chemical Engineering Science*, 16(3-4):242–251, 1961.
- [35] H. A. Lorentz. *Abhandlungen über theoretische Physik*. BG Teubner, 1907.
- [36] D. Sornette. *Critical Phenomena in Natural Sciences (Chaos, Fractals, Self-organization and Disorder: Concepts and Tools)*. Springer Series in Synergetics, Heidelberg, 2004.
- [37] A.W.C. Lau and T. C. Lubensky. State-dependent diffusion: Thermodynamic consistency and its path integral formulation. *Physical Review E*, 76:011123, 2007.
- [38] C. W. Gardiner. *Handbook of stochastic methods*. Springer, 1985.
- [39] J. Ruseckas and B. Kaulakys.  $1/f$  noise from nonlinear stochastic differential equations. *Physical Review E*, 81(3):1–7, 2010.
- [40] P. R. Krugman. Target zones and exchange rate dynamics. *The Quarterly Journal of Economics*, 106(3):669–682, 1991.
- [41] L. E. O. Svensson. Assessing target zone credibility: mean reversion and devaluation expectations in the ERM, 1979–1992. *European Economic Review*, 37(4):763–793, 1993.
- [42] L. E. O. Svensson. The term structure of interest rate differentials in a target zone: Theory and Swedish data. *International Monetary Fund Staff Papers*, 28(1):87–116, 1991.
- [43] G. Bertola and R. J. Caballero. Target zones and realignments. *The American Economic Review*, 82(3):520–536, 1992.
- [44] G. Bertola and L. E. O. Svensson. Stochastic devaluation risk and the empirical fit of target-zone models. *The Review of Economic Studies*, 60(3):689–712, 1993.
- [45] A. M. Werner. Exchange rate target zones, realignments and the interest rate differential: Theory and evidence. *Journal of International Economics*, 39(3):353–367, 1995.
- [46] K. A. Froot and M. Obstfeld. Exchange-rate dynamics under stochastic regime shifts: A unified approach. *Journal of International Economics*, 31(3):203–229, 1991.

- [47] K. M. Dominguez and P. B. Kenen. On the need to allow for the possibility that governments mean what they say: Interpreting the target-zone model of exchange-rate behavior in the light of ems experience. (Working Paper 3670), 1991.
- [48] F. Delgado and B. Dumas. Monetary contracting between central banks and the design of sustainable exchange-rate zones. *Journal of International Economics*, 34(3):201–224, 1993.
- [49] H. Lindberg and P. Söderlind. Intervention policy and mean reversion in exchange rate target zones: The Swedish case. *The Scandinavian Journal of Economics*, 96(4):499–513, 1994.
- [50] P. M. Garber and L. E. O. Svensson. The operation and collapse of fixed exchange rate regimes. In *Handbook of International Economics*, volume III. 1995.
- [51] A. Serrat. Exchange rate dynamics in a multilateral target zone. *Review of Economic Studies*, 67(1):193–211, 2000.
- [52] M. Flandreau. The Burden of Intervention: Externalities in Multilateral Exchange Rates Arrangements. *Journal of International Economics*, 45(1):137–171, 1998.
- [53] H. C. Lin. Forward-rate target zones and exchange rate dynamics. *Journal of International Money and Finance*, 27(5):831–846, 2008.
- [54] L. Bartolini and A. Prati. Soft exchange rate bands and speculative attacks: Theory, and evidence from the ERM since August 1993. *Journal of International Economics*, 49(1):1–29, 1999.
- [55] S. Lundbergh and T. Teräsvirta. A time series model for an exchange rate in a target zone with applications. *Journal of Econometrics*, 131(1-2):579–609, 2006.
- [56] C.-H. Hui and T. Fong. Is the Hong Kong dollar exchange rate bounded in the convertibility zone? *Hong Kong Monetary Authority Working Paper*, (13), 2007.
- [57] M. Funke, Y.-F. Chen, and N. Glanemann. A Soft Edge Target Zone Model: Theory and Application to Hong Kong. *BOFIT Discussion Paper*, (21), 2009.
- [58] L. Sarno and M. Taylor. *The economics of exchange rates*. Cambridge, University Press, 2003.
- [59] L. E. O. Svensson. An interpretation of recent research on exchange rate target zones. *Journal of Economic Perspectives*, 6(4):119–144, 1992.
- [60] A. K. Dixit. *The art of smooth pasting*, volume 55. Taylor & Francis, 1993.
- [61] R. Studer-Suter and A. Janssen. The Swiss franc’s honeymoon. *SSRN*, (Working Paper No. 170), 2014.
- [62] L. E. O. Svensson. The Simplest Test of Target Zone Credibility. *Journal of International Economics*, 38:655–665, 1991.

- [63] R.P. Flood, A.K. Rose, and D.J. Mathieson. An empirical exploration of exchange-rate target-zones. *Carnegie-Rochester Conference Series on Public Policy*, 35:7–65, 1991.
- [64] F. Giavazzi and A. Giovannini. Monetary policy interactions under managed exchange rates. *Economica*, pages 199–213, 1989.
- [65] K. M. Dominguez and P. B. Kenen. Intramarginal intervention in the EMS and the target-zone model of exchange-rate behavior. *European Economic Review*, 36(8):1523–1532, 1992.
- [66] H. Lindberg and P. Söderlind. Testing the basic target zone model on Swedish data 1982-1990. *European Economic Review*, 38(7):1441–1469, 1994.
- [67] R. A. Meese and A. K. Rose. Nonlinear, nonparametric, nonessential exchange rate estimation. *The American Economic Review*, 80(2):192–196, 1990.
- [68] G. W. Smith and M. G. Spencer. Estimation and testing in models of exchange rate target zones and process switching. In P.R. Krugman and M.H. Miller, editors, *Exchange rate targets and currency bands*, pages 211–239. Cambridge University Press, 1992.
- [69] K. G. Koedijk, P. A. Stork, and C. G. De Vries. Differences between foreign exchange rate regimes: the view from the tails. *Journal of International Money and Finance*, 11(5):462–473, 1992.
- [70] M. Iannizzotto and M. P. Taylor. The Target Zone Model, Non-Linearity and Mean Reversion: is the Honeymoon Really Over? *The Economic Journal*, 109(454):96–110, 1999.
- [71] M. P. Taylor and M. Iannizzotto. On the mean-reverting properties of target zone exchange rates: a cautionary note. *Economics Letters*, 71(1):117–129, 2001.
- [72] J. M. Harrison. Brownian motion and stochastic flow systems. *Applied Optics*, 25:31–45, 1986.
- [73] C.F. Lo, C.H. Hui, T. Fong, and S.W. Chu. A quasi-bounded target zone model: Theory and application to hong kong dollar. *International Review of Economics & Finance*, 37:1–17, 2015.
- [74] C. Bauer, P. De Grauwe, and S. Reitz. Exchange rate dynamics in a target zone - A heterogeneous expectations approach. *Journal of Economic Dynamics and Control*, 33(2):329–344, 2009.
- [75] A. Portugal Duarte, J. Sousa Andrade, and A. Duarte. Exchange rate target zones: A survey of the literature. *Journal of Economic Surveys*, 27(2):247–268, 2013.
- [76] A. Naszodi. The options-based model of target zone and its application to the French franc. *Journal of Derivatives & Hedge Funds*, 16(3):212–218, 2010.
- [77] M. Hanke, R. Poulsen, and A. Weissensteiner. Where would the EUR/CHF exchange rate be without the SNB’s minimum exchange rate policy? *Journal of Futures Markets*, 35(12):1103–1116, 2015.

- [78] U. J. Jermann. Financial markets' views about the euro-Swiss franc floor. *Working Paper: National Bureau of Economic Research*, 2016.
- [79] M. Hanke, R. Poulsen, and A. Weissensteiner. Analyzing the Swiss National Bank's euro exchange rate policy: A latent likelihood approach. *SSRN working paper*, (2515210), 2016.
- [80] P. M. C. Morse and H. Feshbach. *Methods of Theoretical Physics*. Technology Press, 1946.
- [81] D. Veestraeten. Currency option pricing in a credible exchange rate target zone. *Applied Financial Economics*, 23(11):951–962, 2013.
- [82] E. Derman and I. Kani. Riding on a smile. *Risk*, 7(2):32–39, 1994.
- [83] R. Van Der Kamp. *Local volatility modelling*. Master thesis, University of Twente, 2009.
- [84] Y. Bergman. General restrictions on prices of financial derivatives written on underlying diffusions. *Working Paper*, 1998.
- [85] Y. Aït-Sahalia. Transition densities for interest rate and other nonlinear diffusions. *Journal of Finance*, 54(4):1361–1395, 1999.
- [86] Y. Aït-Sahalia. Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, 70(1):223–262, 2002.
- [87] L. F. Richardson. The approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application to the stresses in a masonry dam. *Philosophical Transactions of the Royal Society A*, 210:459–470, 1911.
- [88] Hong Kong Monetary Authority, HK. Annual Report, 2005.
- [89] J. C. Cox and S. A. Ross. The valuation of options for alternative stochastic processes. *Journal of financial economics*, 3(1-2):145–166, 1976.
- [90] M.K. Transtrum, B.B. Machta, and J.P. Sethna. Geometry of nonlinear least squares with applications to sloppy models and optimization. *Phys. Rev. E*, 83:673–701, 2011.
- [91] G. Demos and D. Sornette. Birth or burst of financial bubbles: which one is easier to diagnose? *Quantitative Finance*, 17(5):657–675, 2017.
- [92] G.N. Mil'shtejn. Approximate integration of stochastic differential equations. *Theory of Probability & Its Applications*, 19(3):557–562, September 1973.
- [93] L. V. Ballestra and L. Cecere. A numerical method to estimate the parameters of the cev model implied by american option prices: Evidence from nyse. *Chaos, Solitons & Fractals*, 88:100–106, 2016.
- [94] R. Studer-Suter and A. Janssen. The swiss franc's honeymoon. *University of Zurich, Department of Economics, Working Paper*, (170), 2014.
- [95] Y.-K. Kwok. *Mathematical models of financial derivatives*. Springer, 2008.

- [96] C. A. Ball and A. Roma. Target zone modelling and estimation for european monetary system exchange rates. *Journal of Empirical Finance*, 1(3):385–420, 1994.
- [97] D. Veestraeten. Conditional distributions in the Krugman target zone model and undeclared narrower bands. *Center for Economic Studies Discussions Paper Series (DPS) 01.02*, 2001.
- [98] G. Erber. What is unorthodox monetary policy? *SSRN Working Paper*, 2012.
- [99] E. Dabla-Norris, S. Guo, V. Haksar, M. Kim, K. Kochhar, K. Wiseman, and A. Zdzienicka. The new normal: a sector-level perspective on productivity trends in advanced economies. *IMF staff discussion note SDN/15/03*, 2015.
- [100] Y. Louzoun, S. G. Solomon, J. Goldenberg, and D. Mazursky. World-size global markets lead to economic instability. *Artificial life*, 9(4):357–370, 2003.
- [101] D. Sornette and P. Cauwels. 1980-2008: The illusion of the perpetual money machine and what it bodes for the future. *Risks*, 2:103–131, 2014.
- [102] D. Sornette. Nurturing breakthroughs: lessons from complexity theory. *Journal of Economic Interaction and Coordination*, 3(2):165–181, 2008.
- [103] M. Gisler and D. Sornette. Exuberant innovations: the apollo program. *Society*, 46(1):55–68, 2009.
- [104] M. Gisler and D. Sornette. Bubbles everywhere in human affairs. *chapter in book entitled “Modern RISC-Societies. Towards a New Paradigm for Societal Evolution”, L. Kajfez-Bogataj, K. H. Müller, I. Svetlik, N. Tos (eds.), Wien, edition echoraum: 137-153, (<http://ssrn.com/abstract=1590816>), 2010.*
- [105] M. Gisler, D. Sornette, and R. Woodard. Innovation as a social bubble: The example of the human genome project. *Research Policy*, 40(10):1412–1425, 2011.
- [106] J. D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384, 1989.
- [107] J. M. Durland and T. M. McCurdy. Duration-dependent transitions in a Markov model of U.S. GNP growth. *Journal of Business & Economic Statistics*, 12(3):279–288, 1994.
- [108] F. X. Diebold and G. D. Rudebusch. A nonparametric investigation of duration dependence in the american business cycle. *Journal of Political Economy*, 98(3):596–616, 1990.
- [109] V. Filimonov, G. Demos, and D. Sornette. Modified profile likelihood inference and interval forecast of the burst of financial bubbles. *Quantitative Finance*, DOI: 10.1080/14697688.2016.1276298:1–20, 2017.
- [110] R. J. Hodrick and E. C. Prescott. Postwar US business cycles: an empirical investigation. *Journal of Money, credit, and Banking*, pages 1–16, 1997.

- [111] M.. Baxter and R. G. King. Measuring business cycles: Approximate band-pass filters for economic time series. *The Review of Economics and Statistics*, 4(81):575–593, 1999.
- [112] G. H. Moore and V. Zarnowitz. The development and role of the national bureau of economic research’s business cycle chronologie. In *NBER Working Paper*, volume 1394. 1984.
- [113] A. F. Burns and W. C. Mitchell. *Measuring business cycles*. 1946.
- [114] J. Morlet, G. Arens, E. Fourgeau, and D. Glard. Wave propagation and sampling theory-part i: Complex signal and scattering in multilayered media. *Geophysics*, 47(2):203–221, 1982.
- [115] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies. Image coding using wavelet transform. *Image Processing, IEEE Transactions on*, 1(2):205–220, 1992.
- [116] E. Slezak, A. Bijaoui, and G. Mars. Identification of structures from galaxy counts-use of the wavelet transform. *Astronomy and Astrophysics*, 227:301–316, 1990.
- [117] F. Argoul, A. Arneodo, G. Grasseau, and Y. Gagne. Wavelet analysis of turbulence reveals the multifractal nature. *Nature*, 338:2, 1989.
- [118] J. B. Ramsey. The contribution of wavelets to the analysis of economic and financial data. *Philosophical Transactions of the Royal Society A*, 357(1760):2593–2606, 1999.
- [119] Patrick M. Crowley. A guide to wavelets for economists. *Journal of Economic Surveys*, 21(2):207–267, 2007.
- [120] M. Yogo. Measuring business cycles: A wavelet analysis of economic time series. *Economics Letters*, 100(2):208–212, 2008.
- [121] L. Aguiar-Conraria and M. Joana Soares. The continuous wavelet transform: Moving beyond uni- and bivariate analysis. *Journal of Economic Surveys*, 28(2):344–375, 2014.
- [122] D. Ardila and D. Sornette. Dating the financial cycle with uncertainty estimates: a wavelet proposition. *Finance Research Letters*, 19:298–304, 2016.
- [123] P. Yiou, D. Sornette, and M. Ghil. Data-adaptive wavelets and multi-scale singular-spectrum analysis. *Physica D*, 142(3-4):254–290, 2000.
- [124] I. Daubechies. *Ten lectures on wavelets*. SIAM, 1992.
- [125] A. Arneodo, F. Argoul, J.F. Muzy, M. Tabard, and E. Bacry. Beyond classical multifractal analysis using wavelets: Uncovering a multiplicative process hidden in the geometrical complexity of diffusion limited aggregates. *Fractals*, 1(03):629—649, 1993.
- [126] J. G. Fernald and C. I. Jones. The future of US economic growth. *American Economic Review: Papers & Proceedings*, 104(5):44–49, 2014.
- [127] B. W. Silverman. Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society*, 43(1):97–99, 1981.



- [128] P. Geraskin and D. Fantazzini. Everything you always wanted to know about log-periodic power laws for bubble modeling but were afraid to ask. *The European Journal of Finance*, 19(5):366–391, 2013.
- [129] D. Sornette and P. Cauwels. Financial bubbles: mechanisms and diagnostics. *Review of Behavioral Economics*, 2(3):279–305, 2015.
- [130] V.I. Yukalov, E.P. Yukalova, and D. Sornette. Dynamical system theory of periodically collapsing bubbles. *The European Physical Journal B*, 88(7):1–14, 2015.
- [131] T. Atkinson, D. Luttrell, and H. Rosenblum. How bad was it? the costs and consequences of the 2007-09 financial crisis. *Staff Papers Federal Reserve Bank of Dallas*, 20:1–22, 2013.
- [132] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, and D. R. White. Economic networks: The new challenges. *Science*, 325:422–424, 2009.
- [133] X. Gabaix. The granular origins of aggregate fluctuations. *Econometrica*, 79(3):733–772, 2011.
- [134] D. Acemoglu, V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi. The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016, 2012.
- [135] J. Di Giovanni, A. A. Levchenko, and I. Mejean. Firms, destinations, and aggregate fluctuations. *Econometrica*, 82(4):1303–1340, 2014.
- [136] R. Friberg and M. Sanctuary. The contribution of firm-level shocks to aggregate fluctuations: The case of Sweden. *Economics Letters*, 147:8–11, 2016.
- [137] N. Anthonisen. Microeconomic shocks and macroeconomic fluctuations in a dynamic network economy. *Journal of Macroeconomics*, 47(2011):233–254, 2016.
- [138] A. Stella. Firm dynamics and the origins of aggregate fluctuations. *Journal of Economic Dynamics and Control*, 55:71–88, 2015.
- [139] G. Fagiolo, M. Napoletano, and A. Roventini. Are output growth-rate distributions fat-tailed? some evidence from OECD countries. *Journal of Applied Econometrics*, 23(5):639–669, 2008.
- [140] G. Fagiolo, M. Napoletano, M. Piazza, and A. Roventini. Detrending and the distributional properties of US output time series. *Economics Bulletin*, 29(4):3155–3161, 2009.
- [141] D. Fu, F. Pammolli, S. V. Buldyrev, M. Riccaboni, K. Matia, K. Yamasaki, and H. E. Stanley. The growth of business firms: Theoretical framework and empirical evidence. *Proceedings of the National Academy of Sciences of the United States of America*, 102(52):18801–18806, 2005.
- [142] M. A. Williams, G. Baek, Y. Li, L.Y. Park, and W. Zhao. Global evidence on the distribution of GDP growth rates. *Physica A*, 468:750–758, 2017.
- [143] R. Franke. How fat-tailed is US output growth? *Metroeconomica*, 66(2):213–242, 2015.

- [144] Y. Malevergne, V. Pisarenko, and D. Sornette. Empirical distributions of stock returns: Between the stretched exponential and the power law? *Quantitative Finance*, 5(4):379–401, 2005.
- [145] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *Society for Industrial and Applied Mathematics*, 51(4):43, 2007.
- [146] B.V. Gnedenko and A. N. Kolmogorov. *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, 1954.
- [147] C. W. Gardiner. *Stochastic methods: a handbook for the natural and social sciences*. Springer, 2009.
- [148] A. V. Chechkin, J. Klafter, V. Y. Gonchar, and R. Metzler. Bifurcation, bimodality, and finite variance in confined Lévy flights. *Physical Review E*, 67(1):010102, 2003.
- [149] A. V. Chechkin, V. Y. Gonchar, J. Klafter, R. Metzler, and L. V. Tanatarov. Lévy flights in a steep potential well. *Journal of Statistical Physics*, 115(5-6):1505–1535, 2004.
- [150] A. Janicki and A. Weron. *Simulation and chaotic behavior of alpha-stable stochastic processes*. CRC Press, 1993.
- [151] K. G. Wilson. Problems in physics with many scales of length. *Scientific American*, 241(2):140–157, 1979.
- [152] G. De Wit. Firm size distributions: An overview of steady-state distributions resulting from firm dynamics models. *International Journal of Industrial Organization*, 23(5):423–450, 2005.
- [153] E. Zambrano, A. Hernando, A. F. Bariviera, R. Hernando, and A. Plastino. Thermodynamics of firms’ growth. *Journal of The Royal Society Interface*, 12(112):20150789, 2015.
- [154] M. Gertler and S. Girchrist. Monetary policy, business cycles, and the behavior of small manufacturing firms. *The Quarterly Journal of Economics*, 109(2):309–340, 1994.
- [155] S. J. Davis, J. Haltiwanger, and S. Schuh. Small business and job creation: Dissecting the myth and reassessing the facts. *Small business economics*, 8(4):297–315, 1996.
- [156] P. Pagano and F. Schivardi. Firm size distribution and growth. *Scandinavian Journal of Economics*, 105(2):255–274, 2003.
- [157] Y. Malevergne, P. Santa-Clara, and D. Sornette. Professor Zipf goes to Wall Street. *NBER Working Paper No. 15295*, 2009.
- [158] R. Gibrat. *Les inégalités économiques*. Librairie du Recueil Sirey, Paris, 1931.
- [159] E. W. Montroll and M. F. Shlesinger. On  $1/f$  noise and other distributions with long tails. *Proceedings of the National Academy of Sciences of the United States of America*, 79(10):3380–3383, 1982.

- [160] Didier Sornette. *Critical Phenomena in Natural Sciences*. Springer, 2004.
- [161] Y. Malevergne, V. Pisarenko, and D. Sornette. Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. *Physical Review E*, 83(3):036111, 2011.
- [162] E.G.J. Luttmer. On the mechanics of firm growth. *The Review of Economic Studies*, 78(3):1042–1068, 2011.
- [163] R. Axtell, M. Gallegati, and A. Palestrini. Common components in firms’ growth and the sectors scaling puzzle. *Economics Bulletin*, 12:1–8, 2008.
- [164] M. Wyart and J.-P. Bouchaud. Statistical models for company growth. *Physica A*, 326(1):241–255, 2003.
- [165] M. Takayasu, H. Watanabe, and H. Takayasu. Generalised central limit theorems for growth rate distribution of complex systems. *Journal of Statistical Physics*, 155(1):47–71, 2014.
- [166] X. Gabaix. Zipf’s law for cities: an explanation. *Quarterly journal of Economics*, pages 739–767, 1999.
- [167] G. M. Grossman and E. Helpman. *Innovation and growth in the global economy*. MIT press, 1993.
- [168] J. Steindl. *Random processes and the growth of firms: A study of the Pareto law*. Griffin London, 1965.
- [169] E.G.J. Luttmer. Selection, growth, and the size distribution of firms. *The Quarterly Journal of Economics*, 122(3):1103–1144, 2007.
- [170] A. Saichev, Y. Malevergne, and D. Sornette. *Theory of Zipf’s Law and Beyond*. Springer, 2010.
- [171] Y. Malevergne, A. Saichev, and D. Sornette. Zipf’s law and maximum sustainable growth. *Journal of Economic Dynamics and Control*, 37(6):1195–1212, 2013.
- [172] L.M.B. Cabral and J. Mata. On the evolution of the firm size distribution: Facts and theory. *American economic review*, pages 1075–1090, 2003.
- [173] X. Gabaix, J.-M. Lasry, P.-L. Lions, and B. Moll. The dynamics of inequality. Technical report, National Bureau of Economic Research, 2015.
- [174] John Sutton. Gibrat’s legacy. *Journal of economic literature*, 35(1):40–59, 1997.
- [175] A. Segarra and M. Teruel. An appraisal of firm size distribution: Does sample size matter? *Journal of Economic Behavior & Organization*, 82(1):314–328, 2012.
- [176] H. Schenk. *Multidisciplinary Economics: The Birth of a New Economics Faculty in the Netherlands*, chapter Organisational economics in an age of restructuring, or: how corporate strategies can harm your economy, pages 333–365. Springer US, 2005.

- [177] J. Harford. What drives merger waves? *Journal of financial economics*, 77(3):529–560, 2005.
- [178] C. Kummer and U. Steger. Why merger and acquisition (m&a) waves reoccur: the vicious circle from pressure to failure. *Strategic Management Review*, 2(1):44–63, 2008.
- [179] Y. Ijiri and H. A. Simon. Interpretations of departures from the Pareto curve firm-size distributions. *Journal of Political Economy*, 82(2):315, 1974.
- [180] A. Singh. Take-overs, economic natural selections, and the theory of the firm: Evidence from the postwar united kingdom experience. *The Economic Journal*, 85(339):497, 1975.
- [181] S. Aaronovitch and M.C. Sawyer. *Big business : theoretical and empirical aspects of concentration and mergers in the United Kingdom*. London: Macmillan, 1975.
- [182] Y. Ijiri and H. A. Simon. Effects of mergers and acquisitions on business firm concentration. *Journal of Political Economy*, 79(2):314, 1971.
- [183] E. Cefis, O. Marsili, and H. Schenk. The effects of mergers and acquisitions on the firm size distribution. *Journal of Evolutionary Economics*, 19(1):1–20, 2009.
- [184] L. Hannah and J. A. Kay. *Concentration in modern industry: Theory, measurement and the UK experience*. London: Macmillan, 1977.
- [185] P.E. Hart. On bias and concentration. *The Journal of Industrial Economics*, pages 211–226, 1979.
- [186] L. Hannah and J. A. Kay. The contribution of mergers to concentration growth: a reply to professor hart. *The Journal of Industrial Economics*, pages 305–313, 1981.
- [187] M. Smoluchowski. Drei vorträge über diffusion, brownsche molekularebewegung und koagulation von kolloidteilchen. *Zeitschrift für Physik*, 17(1):557–571, 587–599, 1916.
- [188] M. Smoluchowski. Versuch einer mathematischen theorie der koagulationskinetik kolloider lösungen. *Zeitschrift für physikalische Chemie*, 92(2):129–168, 1917.
- [189] D. J. Aldous. Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. *Bernoulli*, 5(1):3–48, 1999.
- [190] F. Leyvraz. Scaling theory and exactly solved models in the kinetics of irreversible aggregation. *Physics Reports*, 383(2-3):95–212, 2003.
- [191] J. A. D. Wattis. An introduction to mathematical models of coagulation–fragmentation processes: a discrete deterministic mean-field approach. *Physica D*, 222(1):1–20, 2006.
- [192] N. Fournier and P. Lauren. Physics Existence of Self-Similar Solutions to Smoluchowski’s Coagulation Equation. *Communications in Mathematical Physics*, 256:589–609, 2005.
- [193] Z. Mimouni and J. A. D. Wattis. Similarity solution of coagulation equation with an inverse kernel. *Physica A*, 388(7):1067–1073, 2009.

- [194] B. Niethammer, S. Throm, and J. J. L. Velázquez. Self-similar solutions with fat tails for Smoluchowski's coagulation equation with singular kernels. *Communications in Mathematical Physics*, 318(2):505–532, 2013.
- [195] B. Niethammer and J. J. L. Velázquez. Self-similar solutions with fat tails for Smoluchowski's coagulation equation with locally bounded kernels. *Communications in Mathematical Physics*, 318(2):505–532, 2013.
- [196] F.P. Da Costa. Mathematical aspects of coagulation-fragmentation equations. In *Mathematics of Energy and Climate Change*, pages 83—162. Springer, 2015.
- [197] A.I. Saichev and G.M. Zaslavsky. Fractional kinetic equations: solutions and applications. *Chaos*, 7(4):753–764, 1997.
- [198] O. N. Repin and A. I. Saichev. Fractional poisson law. *Radiophysics and Quantum Electronics*, 43(9):738–741, 2000.
- [199] N. Laskin. Fractional poisson process. *Communications in Nonlinear Science and Numerical Simulation*, 8(3-4):201–213, 2003.
- [200] E.M. Hendriks, M.H. Ernst, and R. M. Ziff. Coagulation equations with gelation. *Journal of Statistical Physics*, 31(3):519–563, 1983.
- [201] A. D. Myshkis. Mixed functional differential equations. *Journal of Mathematical Sciences*, 129(5):4111–4226, 2005.
- [202] M. H. Lee. On the validity of the coagulation equation and the nature of runaway growth. *Icarus*, 143:74–86, 2000.
- [203] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events: for insurance and finance*. Springer Science & Business Media, 2013.
- [204] G. Menon and R. L. Pego. Approach to self-similarity in Smoluchowski's coagulation equations. *Communications on pure and applied mathematics*, 57(9):1197–1232, 2004.
- [205] G. Menon and R. L. Pego. Dynamical scaling in Smoluchowski's coagulation equations: uniform convergence. *SIAM review*, 48(4):745–768, 2006.
- [206] H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440, 1955.
- [207] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [208] G. Bianconi and A. L. Barabasi. Competition and multiscaling in evolving networks. *Europhysics Letters*, 54(4):436–442, 2001.
- [209] M. Bell, S. Perera, M. Piraveenan, M. Bliemer, T. Latty, and C. Reid. Network growth models: A behavioural basis for attachment proportional to fitness. *Scientific Reports*, 7, 2017.

- [210] G. Bianconi and A. L. Barabasi. Bose-Einstein condensation in complex networks. *Physical Review Letters*, 86(24):5632–5, 2001.
- [211] S.N. Dorogovtsev and J.F.F. Mendes. Scaling properties of scale-free evolving networks: continuous approach. *Physical Review E*, 63:056125, 2001.
- [212] P. L. Krapivsky and S. Redner. A statistical physics perspective on web growth. *Computer Networks*, 39(3):261–276, 2002.
- [213] Didier Sornette. Dragon-kings, black swans, and the prediction of crises. *International Journal of Terraspace Science and Engineering*, 2(1):1–18, 2009.
- [214] D. Sornette and G. Ouillon. Dragon-kings: Mechanisms, statistical methods and empirical evidence. *The European Physical Journal Special Topics*, 205(1):1–26, 2012.
- [215] J. S. Kong, N. Sarshar, and V. P. Roychowdhury. Experience versus talent shapes the structure of the web. *Proceedings of the National Academy of Sciences*, 105(37):13724–13729, 2008.
- [216] Joseph S. Kong and Vwani P. Roychowdhury. Preferential survival in models of complex ad hoc networks. *Physica A*, 387(13):3335–3347, 2008.
- [217] S. Dereich and P. Mörters. Random networks with sublinear preferential attachment: the giant component. *The Annals of Probability*, 41(1):329–384, 2013.
- [218] M. E. J. Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.
- [219] N. Sarshar and V. Roychowdhury. Scale-free and stable structures in complex ad hoc networks. *Physical Review E*, 69(2 2):1–6, 2004.