NELSON DEUS RODRIGUES

APPLICATION OF GAUGE THEORY TO FINANCE: A SYSTEMATIC LITERATURE REVIEW

UNIVERSITY OF ALGARVE

Faculty of Economics

2019

NELSON DEUS RODRIGUES

APPLICATION OF GAUGE THEORY TO FINANCE:
A SYSTEMATIC LITERATURE REVIEW

Master in Financial Economics

Dissertation made under the supervision of:

Prof. Pedro Miguel Guerreiro Patolea Pintassilgo
Prof. Cristina Maria Pereira Viegas de Oliveira

UNIVERSITY OF ALGARVE

Faculty of Economics

2019

APPLICATION OF GAUGE THEORY TO FINANCE:

A SYSTEMATIC LITERATURE REVIEW

Work Authorship Declaration

I declare to be the author of this work, which is unique and unprecedented. Authors and works consulted are properly cited in the text and are included in the listing of references.

NELSON DEUS RODRIGUES

Copyright

ACKNOWLEDGMENTS

I would like to manifest great gratitude towards those who've helped me in the creation of this dissertation.

I'm deeply thankful to have had Professors Pedro Pintassilgo and Cristina Oliveira as my supervisors, who have given me constant support and helpful advice throughout this work, always with valuable suggestions and information to share.

I would like to thank Professors Luís Coelho, Rúben Peixinho and Nenad Manojlović, for providing feedback regarding some aspects of this work. A special thanks in particular to Professor Nenad, for also helping me with the more mathematical aspects of this dissertation.

I would also like to thank all my friends for their help and affection, and a special thanks to my friends José Diogo Simão, for providing me with mathematical references and discussions about differential geometry, and Ricardo Candeias, for his assistance in the creation of the illustrations that appear in this dissertation.

Finally, it would have been impossible for any of this to come to fruition without the love, affection and encouragement of my family. Therefore, I would like to thank my grandparents and aunts, who've demonstrated unconditional support throughout this dissertation, and especially my mother, father and sister, for whom I am profoundly grateful to have in my life.

ABSTRACT

In this dissertation, a systematic literature review was undertaken, exploring the application of gauge theory, an important formalism in physics literature, to finance. A set of keywords pertaining both gauge theory and finance were established and used as a search string in the database Web of Science. After exclusion and inclusion principles were applied to the set of articles generated, 14 papers were obtained.

By systematically reviewing them, three major approaches to a financial gauge theory were found: Beliefs-Preferences Gauge Symmetry, Local *Numéraire* Gauge Symmetry, and Deflator-Term Structure Gauge Symmetry. These can be essentially differentiated by the kind of gauge symmetry explored. Changing pairs of beliefs and preferences, local *numéraires* and pairs of deflator and term structure is argued to be of no consequence to the dynamics of the financial market under consideration. A differential geometric treatment of financial markets as fibre bundles was shown to be necessary for an understanding of the gauge theory application, and proved itself to be successful in rethinking certain concepts, such as gains from arbitrage opportunities, being equivalent to the curvature of the said fibre bundle, an invariant under gauge transformations.

The local *numéraire* gauge symmetry turned out to be the most investigated one, leading to the execution of various numerical simulations, each with different added variations. Amongst them, the idea of using path integrals, a formalism from quantum mechanics, as a way of simulating the log price probability distributions of a market is used. This works by assuming that the market is characterized by the minimization of arbitrage opportunities. It was found good agreement with historical data, which substantiates the existence of gauge symmetry in financial markets, at least to some extent.

Keywords: Gauge Theory; Finance; Systematic Literature Review; Differential Geometry

RESUMO

Nesta dissertação de mestrado foi realizada uma revisão sistemática da literatura, visando investigar a aplicação de teorias de *gauge* no contexto financeiro. Para este fim, foi construído um conjunto de palavras-chave, pertinentes tanto em finanças como em teorias de *gauge*, subsequentemente introduzidas na base de dados *Web of Science*, com o intuito de encontrar todos os artigos que de alguma maneira as abordem. Princípios de exclusão e inclusão foram aplicados ao conjunto de artigos previamente obtido, traduzindo-se em 14 artigos considerados pertinentes. Revendo-os de modo sistemático, conclui-se que três abordagens para uma teoria de *gauge* financeira podem ser destiladas: Simetria de *Gauge* Crenças-Preferências, Simetria de *Gauge Numéraire* local, e Simetria de *Gauge* Deflator-Termo de Estrutura. Estas diferem no género de simetria de *gauge* explorada. Alterações que afectem pares de crenças e preferências, *numéraire* locais e pares de deflatores e termos de estrutura, assumem-se de nenhuma consequência no que diz respeito às dinâmicas do mercado financeiro em consideração. Para um entendimento de teorias de gauge, provou-se necessário um tratamento geométrico de mercados financeiros, inspirado pelo formalismo de geometria diferencial, interpretando-os como um feixe de fibras. Tal tratamento matemático permite a reconceptualização de ganhos ocorridos por usufruir de oportunidades de arbitragem como elementos do tensor de curvatura do feixe de fibras. Esta quantidade é dita invariante perante transformações de *gauge*.

A simetria de *gauge* associada a escolhas locais de *numéraire* revelou-se a abordagem mais investigada, levando à execução de diversas simulações numéricas, cada uma com adições únicas ao modelo base. A ideia de usar integrais de caminho, um formalismo comum em mecânica quântica, de maneira a simular as distribuições de probabilidades do logaritmo de preços característicos de um mercado financeiro serviu de modelo base. Tal modelo baseia-se na assunção de que um mercado financeiro é caracterizado por minimizar os possíveis ganhos associados a oportunidades de arbitragem. Demonstrou-se uma boa concordância entre dados históricos, relativo aos preços de diversos activos, substanciando a ideia fundamental de que simetria de *gauge* existe em mercados financeiros.

Keywords: Teoria de Gauge; Finanças; Revisão Sistemática da Literatura; Geometria Diferencial

FIGURES INDEX

TABLES INDEX

GENERAL INDEX

1 INTRODUCTION

Financial markets constitute complex systems whose dynamics have been best described and modelled with the aid of tools borrowed from mathematics, and, to some extent, physics (Ilinski, 2000; Smolin, 2009). Such an influence from the deemed "exact sciences" onto financial and economic theory could be potentially attributed to the nature of problem-solving found in these disciplines. A major difference between "exact sciences" and financial economics is the human element at play, introducing an additional difficulty, due to its unpredictable nature (Smolin, 2009). However, as with any model describing reality, whether it's physical, mathematical or human reality, assumptions regarding their nature must always be established. Any naive hopes of total and unequivocal description of nature must be abandoned, regardless of the field. For this reason, mathematics applied to human nature should not be considered an impairing oversimplification, but a development of models which attempt to describe an unpredictable nature, given some set of axioms, hopefully, sufficiently realistic. One may criticize the assumptions, deeming them as potentially reductive, but one cannot criticize (assuming that a correct procedure was undertaken during the research) the conclusions derived.

Many examples of the applicability of mathematics and physics can be found throughout financial literature: the seemingly random motion of small gas particles (but large enough to be observed under a microscope) and Bachelier's model of stock market prices are both described as Brownian motions (Ilinski, 2001); Mantegna & Stanley (2000) with their ideas of scale invariance, and other topics pervasive in physics literature, leading to interdisciplinary subjects such as econophysics and mathematical finance; applications of quantum mechanics to option pricing, by employing path integrals (to be briefly discussed in subsection 5.2.1) and other formalisms (Baaquie, 2004) and also the realization that the famous Black-Scholes equation of option pricing theory bears undeniable resemblance with the Fokker-Planck equation in statistical physics (Young, 1999).

In the realm of economic theory, one finds the development of game theory and general equilibrium theory, whose highly mathematical mechanisms allowed for the establishment of a sufficiently strong bedrock on which many aspects of economic theory rest (Smolin, 2009).

Following this "tradition", we here manifest an interest of assessing in what ways has gauge theory been adapted to financial economics and what kind of perks are attained from adopting such a formalism. Gauge theory can be briefly defined as any theory which describes the dynamics of a system, under the constraint that its dynamics must be invariant under some transformations, considered to be resultant from arbitrary observer's choice (Trautman, 1981; Smolin, 2009). It can also be thought of as a system whose dynamics are endowed with extra degrees of freedom. Gauge theoretical concepts have proven themselves of paramount importance in a great variety of sub-fields in physics, appearing in a great number of fundamental theories e.g. Quantum Electrodynamics, General Relativity, among others (François et al., 2014). In that context, a suggestive reason to such symmetry being so important is the fundamental idea that the laws that describe physical nature should be independent of arbitrary conventions.

In this dissertation, a systematic literature review is performed, in order to understand to what extent gauge theory found application to finance. To assure the replicability of the review, the methodology found in Tranfield et al. (2003) was followed. We aim to assess the usefulness of a gauge theoretical approach to finance by critically reviewing the selected articles and exposing existing gaps in the literature.

In a financial scenario, it's expected that dynamics remain invariant under certain changes regarding arbitrary conventions. For this situation, the entities whose dynamics we are interested are prices of assets, whether these are shares, bonds, currencies, indices, financial derivatives or others. For the exposition of the application of gauge theory to finance, it became necessary to acquire a basic grasp of the formalism of gauge theory. With that in mind, an understanding of the underlying mathematics proves itself indispensable, which belong to the quite dense subfield of differential geometry. For this reason, a chapter of the dissertation is dedicated to a heuristic exposition of the mathematical formalism.

The dissertation is organized as follows: in Chapter 2, an exposition of the mathematical formalism with which gauge theories are best understood is given; Chapter 3 describes the methodology used to perform the systematic literature review, while Chapter 4 shows the obtained results, such as the papers encountered and brief descriptive accounts; Chapter 5 provides an extensive discussion of the obtained results, by synthesizing all papers, and pointing out the links between them; we end this dissertation with Chapter 6, presenting concluding remarks, such as positive and negative aspects of the found applications of gauge theory to finance, possibilities for future research in this topic, and a critic look onto the utilized methodology.

2 ELEMENTS OF DIFFERENTIAL GEOMETRY

The language on which gauge theory is built relies heavily on a mathematical formalism known as differential geometry (François et al., 2014). Thus, it becomes imperative to possess a basic understanding of differential geometry in order to proceed. Due to the high complexity and vastness of this field of mathematics, we'll present only a few elements of the formalism, just enough so that the language is somewhat understood and the advantages are identified, in order to commence our discussion on gauge theory. We won't aim for a deeply formal account of the theory. Instead, the intuition behind every notion is seen as paramount. Any theorem or proposition exhibited in this summary will not have an associated proof. For these, we point out literature on which we've based every statement.

An important caveat regarding the nature of the discussion to come needs to be addressed. Differential geometry requires elements of point-set topology as bedrock for a proper exposition. However, we won't dedicate any section for this. It is our hope that the intuition behind differential geometry be present, even without this pre-requisite. Any topological notion in need of introduction, will be defined and exemplified immediately, allowing for the preservation of intuitiveness.

2.1 Motivation[1]

We'll begin by considering the standard Euclidean space:

$$\mathbb{R}^n = \{ (x^1, x^2, ..., x^n) \big| x^i \in \mathbb{R}, \forall i \} \tag{2.1}$$

As per usual of Euclidean spaces, they possess global coordinates. That is, regardless of where we are in the space, we can always describe it with the same coordinate system. The standard coordinates would be the Cartesian ones, but it could be many others, such as polar or spherical ones. However, the choice of coordinates is largely arbitrary, depending ultimately on preferences or practicality, and therefore, any concepts which depend on coordinates can't bear the same level of fundamentality. Another way of approaching this idea of coordinate-free importance is by taking a more practical stance. All these coordinates mentioned are nothing more than a way to perform algebraic operations. Nothing more than "human conventions". Any "real" concepts must not depend on conventions. This notion will be important to understand the applicability to finance.

Lastly, when one deals with more complicated geometries, which will be discussed later, global coordinates might not exist, which means that concepts which depend on these

[1] Subsection 2.1 and 2.2 follow references such as Lee (2000), Tu (2011) and Tu (2017).

render themselves useless. A coordinate-free description is necessary.

2.2 Differentiable Manifolds

A n-dimensional, connected, Hausdorff topological space [2] M is a manifold if (Ilinski, 2001; Isham, 1999):

- There exists a family of open subsets $\{U_\alpha | U_\alpha \subset M, \forall\, \alpha\}$ that covers M, that is:

$$M = \bigcup_\alpha U_\alpha \tag{2.2}$$

- There exists homeomorphic maps ϕ_α (that is, continuous maps that "take" each point in the domain to a single point in the image, whose inverse is also continuous), such that:

$$\begin{aligned} \phi_\alpha : U_\alpha &\longrightarrow \phi_\alpha(U_\alpha) \subset \mathbb{R}^n \\ p &\longmapsto \phi_\alpha(p) = (x^1, x^2, ..., x^n) \end{aligned} \tag{2.3}$$

  where $(x^1, x^2, ..., x^n)$ is a point in $\mathbb{R}^n$ in some coordinates. For any two overlapping U_α and U_β, that is, $U_\alpha \cap U_\beta \neq \emptyset$, the following maps are also homeomorphisms:

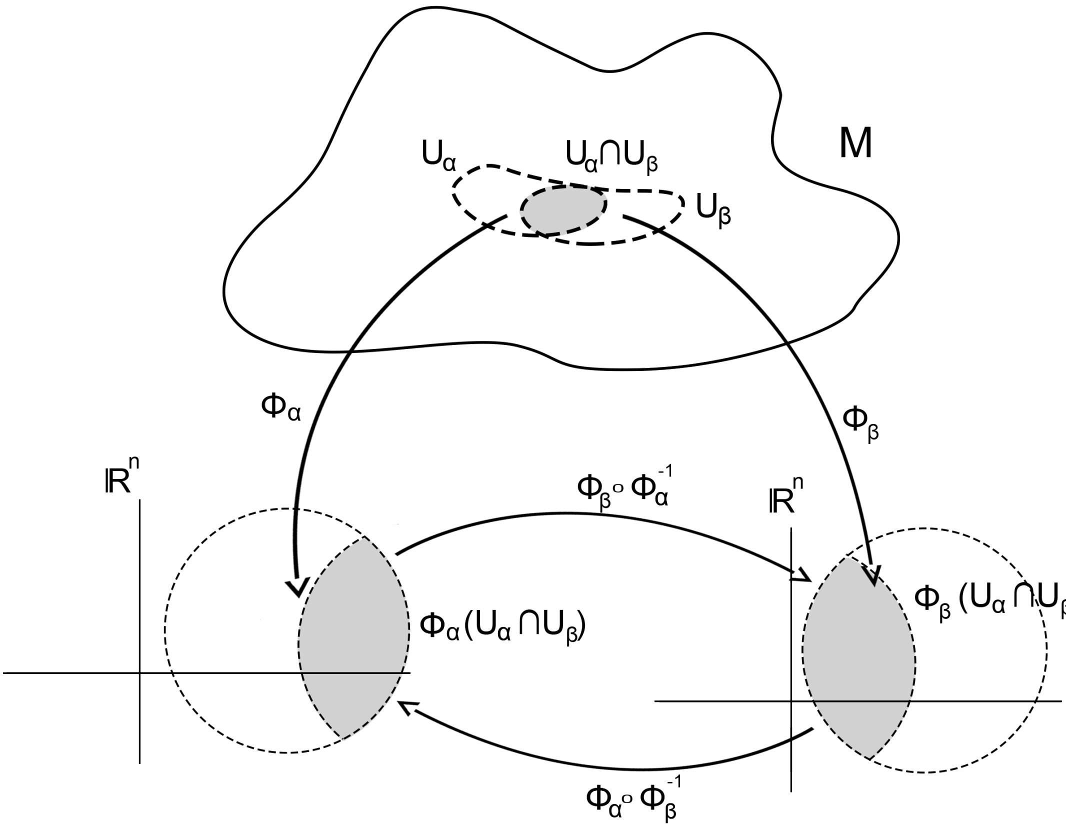
$$\begin{cases} \phi_\alpha \circ \phi_\beta^{-1} : \phi_\beta(U_\alpha \cap U_\beta) \longrightarrow \phi_\alpha(U_\alpha \cap U_\beta) \\ \phi_\beta \circ \phi_\alpha^{-1} : \phi_\alpha(U_\alpha \cap U_\beta) \longrightarrow \phi_\beta(U_\alpha \cap U_\beta) \end{cases} \tag{2.4}$$

  where $\phi_\alpha(U_\alpha \cap U_\beta),\ \phi_\beta(U_\alpha \cap U_\beta) \subset \mathbb{R}^n$. These maps are known as overlap functions.

The pair (U_α, ϕ_α) is known as a local chart, and the collection $\{(U_\alpha, \phi_\alpha) | U_\alpha \subset M, \forall \alpha\}$ is an atlas on M. When the overlap functions in (2.4) are infinitely differentiable (said to be smooth, or of class C^∞), we say that the manifold is a smooth manifold.

[2] See Munkres (2000) for definitions and explanations of these topological terms.

Figure 2.1: Representation of smooth charts on manifold M.



Source: Adapted from Tu (2011).

Intuitively speaking (see Figure 2.1)[3], a n-dimensional smooth manifold is a complicated geometric structure when globally considered, which can be seen as a series of patches in Euclidean space (locally homeomorphic to n-dimensional Euclidean space via a smooth map ϕ_α) all glued together to create the whole object (the union of all subsets U_α covers the manifold, and they're consistent in overlapping regions, see equation (2.4)).

It is this construction that allows to actually work with manifolds from an analytical standpoint. We've generalized to a point where global coordinates no longer exist, since the considered object doesn't "live" in the Euclidean ambient space, losing the ability to effectively analyse it with the mathematical tools from calculus, but then recuperate these by saying that the whole structure can be locally approximated to the Euclidean space in a smooth fashion.

This notion is at the heart of differential geometry: the back and forth between the manifold structure, which is quite general but too abstract, rendering it unworkable with, and the Euclidean space, which is far too specific, but allows standard calculus. All

[3] From now on, any pictorial representation of the Euclidean space in n dimensions will be represented by $\mathbb{R}^2$.

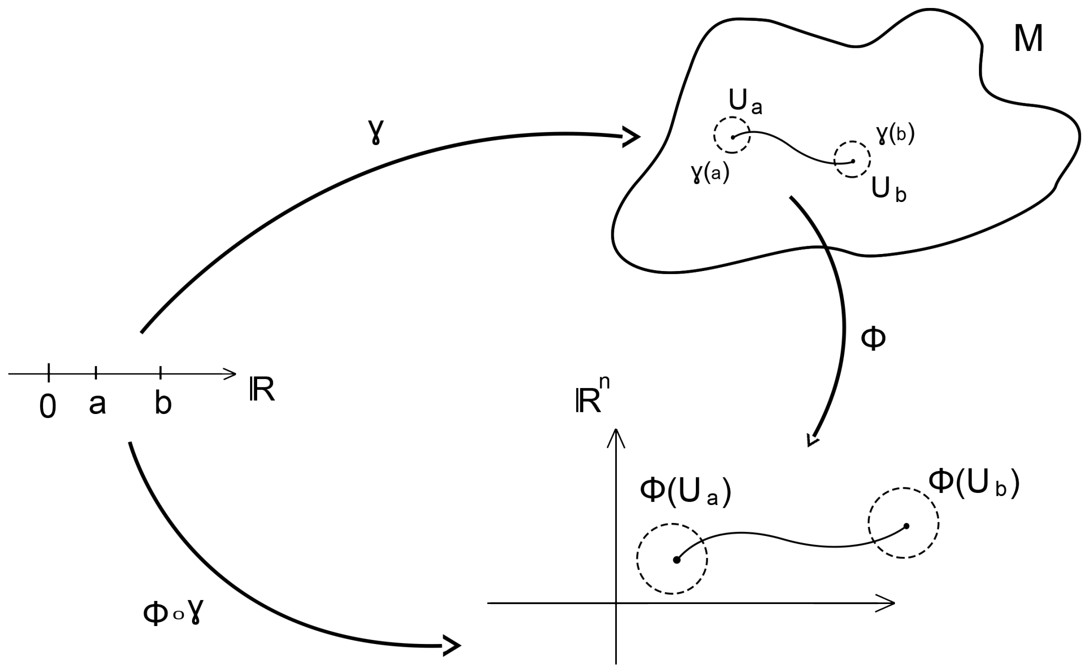
manifolds considered from now on will be assumed to be smooth, unless otherwise stated. Standard tools of calculus can be dealt with in the manifold by recurring back to the Euclidean space, as previously mentioned, but now in a coordinate-free manner. Three notions will be highlighted due to their importance: smooth curves, functions and maps between manifolds.

Let $[a,b] \subset \mathbb{R}$ be a closed interval. A curve γ on a manifold M is defined as the injective map $\gamma : [a,b] \to M$ (see illustration in Figure 2.2). At each point $\gamma(t) \in M$, $\forall t \in [a,b]$, there exists a chart (U_t, ϕ_t). Therefore, at some t, we have the function:

$$\begin{aligned} \phi_t \circ \gamma : [a,b] &\longrightarrow \phi_t(U_t) \subset \mathbb{R}^n \\ t &\longmapsto \phi_t(\gamma(t)) \end{aligned} \tag{2.5}$$

which is a standard function from $\mathbb{R}$ to $\mathbb{R}^2$ that can be evaluated via calculus.

Figure 2.2: Smooth curve on a manifold.



Source: Adapted from Tu (2011).

A function f on M is denoted as $f : M \to \mathbb{R}^k$, where k is not necessarily equal to n. This function is smooth if, for all points $p \in M$ there's a chart (U, ϕ), where $p \in U \subset M$, such that $f \circ \phi^{-1} : \phi(U) \subset \mathbb{R}^n \to \mathbb{R}^k$.

And finally, the notion of a map between manifolds. Let M^n and N^k be n and k dimensional manifolds, respectively. Let there be a bijective map $F : M \to N$ between them. The map is said to be smooth if for all points $p \in M$ there is a chart (U, ϕ), where $p \in U \subset M$, and a chart (V, ψ), where $F(p) \in V \subset N$, such that $F(U) \subset V$ and $\hat{F} \equiv \psi \circ F \circ \phi^{-1} : \phi(U) \subset \mathbb{R}^n \to \mathbb{R}^k \supset \psi(V)$ is smooth.

In summary, despite the fact that the previous functions have a domain or an image which is a general manifold, in practice, we consider them locally, where they are identical with Euclidean space, where, once again, calculus is readily applicable.

2.2.1 Tangent Space

We've seen that a manifold structure can be approximated by the union of an arbitrary number of subsets which are essentially "equal" to the Euclidean space. Here we introduce the idea of tangent space, which can be understood intuitively as a linear approximation at some point of a given surface or a more abstract geometric object. For the case of Euclidean space, we denote the space whose elements are tangent vectors at a some point $p \in \mathbb{R}^n$ as the tangent space $T_p\mathbb{R}^n$. Every manifold is locally Euclidean, therefore $U \subset M$ also has an associated tangent space T_pU. Since the tangent space of a manifold is interpreted as a linear approximation at some point, T_pU and T_pM must be identical. From now on we'll denote the tangent space at a point $p \in M$ as T_pM. A vector $X_p \in T_pM$ can be written in local coordinates $(x^1, x^2, ..., x^n)$ as a linear combination of basis vectors, which can be proven to be the differential operators $\partial/\partial x_i \equiv \partial_i$, $\forall i$ (see, for example, Tu (2011), for such a proof).

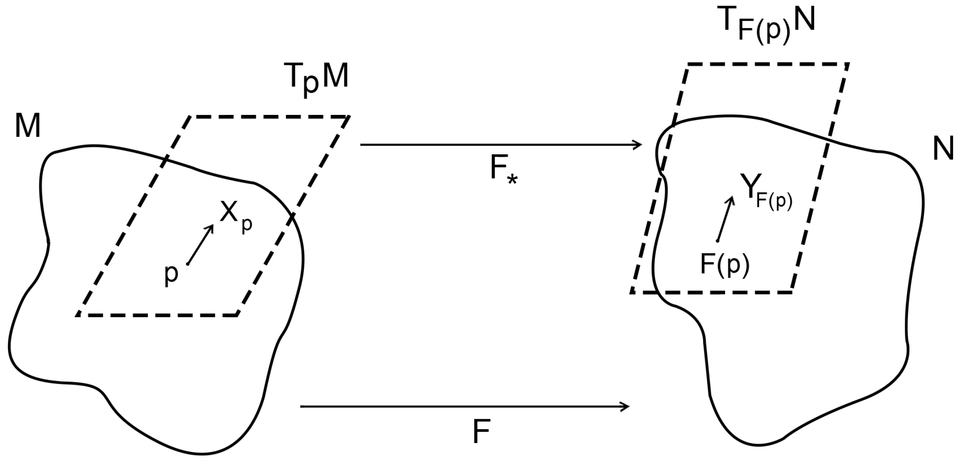
An important map between the tangent spaces of two manifolds is now discussed. Again, let M and N be smooth manifolds, whose dimensions aren't necessarily equal. For each $p \in M$, we define the push-forward of F as:

$$\begin{aligned} F_* : T_pM &\longrightarrow T_{F(p)}N \\ X_p &\longmapsto Y_{F(p)} = F_* X_p \end{aligned} \tag{2.6}$$

where X_p and $Y_{F(p)}$ are tangent vectors of M and N, respectively. Thus, the push-forward provides a rule for mapping tangent vectors between manifolds.

As seen in Figure 2.3, a smooth map F between these two manifolds provides a map between their tangent spaces at each point, the push-forward F_*. This way, vectors between manifolds can be "compared" with one another.

Figure 2.3: Smooth map F and induced push-forward F_*.



Source: Adapted from Tu (2010).

2.2.2 Lie Group

We end this section by introducing an important kind of manifold, which will bear great importance in later sections – the Lie group (any mathematical reference pertinent to differential geometry, usually presents a section dealing with this subject; we point out, for example, Isham (1999), Lee (2000) or Tu (2011)). A Lie group G is a special kind of smooth manifold, endowed with the additional structure of a group, where the following smooth maps are obeyed:

- $\mu : G \times G \to G$ by $(g_1, g_2) \mapsto \mu(g_1, g_2) = g_1 \cdot g_2$ (multiplication)
- $i : G \to G$ by $g \mapsto i(g) = g^{-1}$ (inversion)

An important element of a Lie group is known as the identity e. For any $g \in G$, we have $\mu(e, g) = e \cdot g = g$ and $i(e) = e$.

Like with any other manifold, we can conceptualize a tangent space at some point. The point of interest will be the identity, because, due to the group structure, we can achieve any other point in the manifold by multiplying the identity with some group element. The tangent space of the Lie group at the identity T_eG bears the name of Lie algebra, $\mathscr{G}$, and it comes with a canonical map known as Lie brackets:

$$\begin{aligned} [\cdot,\cdot] : \mathscr{G} \times \mathscr{G} &\longrightarrow \mathscr{G} \\ (X_e, Y_e) &\longmapsto [X_e, Y_e] = X_eY_e - Y_eX_e \end{aligned} \tag{2.7}$$

where X_e and Y_e are two tangent vectors of the Lie group at the identity.

Finally, we point out that from now on, unless stated otherwise, only Abelian groups will be considered. Abelian groups are commutative, that is, $a \cdot b = b \cdot a$, for $a,\ b \in G$. Consequently, the Lie brackets above will always be zero. However, we'll still respect order of application of operators as if they didn't commute, just to keep in accordance with the literature.

2.3 Fibre Bundles

Now we introduce the geometric structure of a fibre bundle, which will play a fundamental role in gauge theoretical language. For this subsection, we follow mostly Ilinski (2001) and Way (2010). A fibre bundle (see Figure 2.4) is a geometric construction with the following constituents:

1. Manifold E called Total Space;

2. Manifold M called Base Space;

3. Manifold F called Fibre Space;

4. Projection map $\pi : E \to M$ (assumed to be surjective) and local trivialization of E (assumed to be a smooth one-to-one map) given by:

$$\begin{aligned} \varphi_\alpha : E \supset \pi^{-1}(U_\alpha) &\longrightarrow (\varphi_\alpha \circ \pi^{-1})(U_\alpha) = U_\alpha \times F \\ \pi^{-1}(p) &\longmapsto (\varphi_\alpha \circ \pi^{-1})(p) = \{p\} \times F_p \simeq F_p \end{aligned} \tag{2.8}$$

   where $F_p = \{f_p \in F \ \ given \ \ p \in U_\alpha\}$ is a set of fibre elements belonging to F, parametrized by $p \in U_\alpha \subset M$. The total space is given by the union of all fibre elements:

$$E = \bigcup_{p \in M} \pi^{-1}(p) = \bigcup_{p \in M} F_p \tag{2.9}$$

   An element of the manifold E will be denoted as the pair $\tilde{p} = (p, f_p)$, which belongs to $F_p \leftrightarrow \pi^{-1}(p)$ and, if projected to the base space, is equal to p;

5. Structural group G (not necessarily a Lie group) acting on the right on the total space E, via smooth transformations:

$$\begin{aligned} E \times G &\longrightarrow E \\ (\tilde{p}, g) &\longmapsto \tilde{p} \cdot g \end{aligned} \tag{2.10}$$

which induces a right action on E:

$$\begin{aligned} R_g : E &\longrightarrow E \\ \tilde{p} &\longmapsto R_g(\tilde{p}) = \tilde{p} \cdot g \end{aligned} \tag{2.11}$$

Figure 2.4: Representation of a fibre bundle geometry.

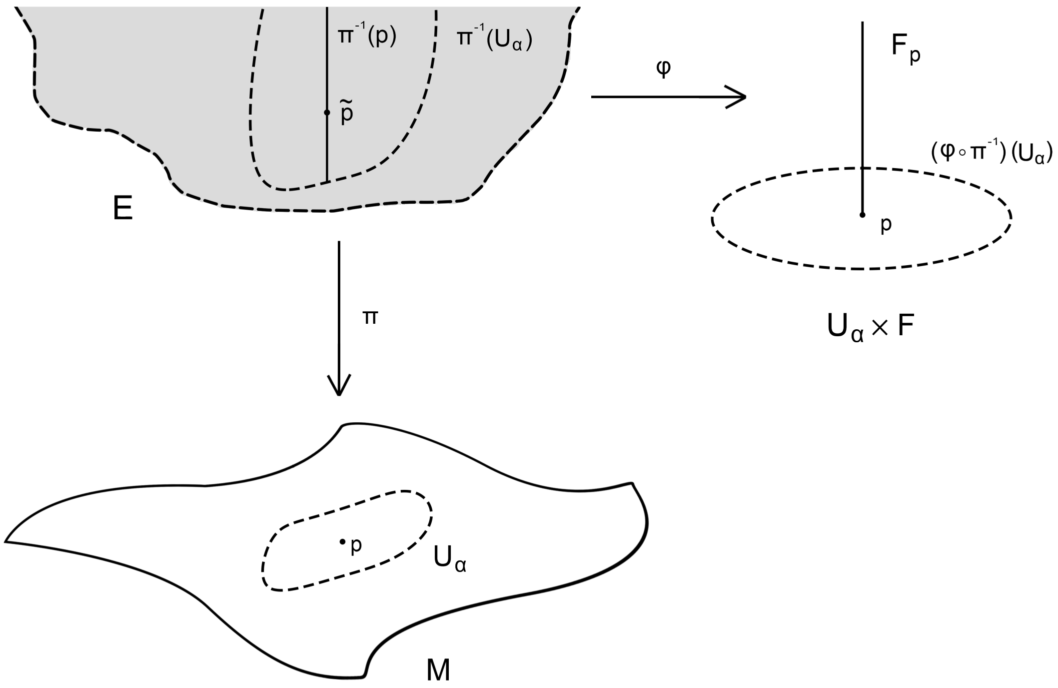


Figure 2.4 is interpreted as follows: to each subset U in the base space M, there's a "correspondent" subset in the total space E, which is locally homeomorphic to the product $F \times U$, where F is the fibre, here represented as lines [4].

What is paramount here is the notion of a manifold M (our "real" space, that is, the space where we actually work on) being the projection of this "complicated" manifold E, which locally can be approximated to a simple Cartesian product between an open subset U_α of M and the fibre F.

2.3.1 Transition Functions

Returning to our previous discussion, now we consider overlapping regions in the base space (see Figure 2.5). For $p \in U_\alpha \cap U_\beta$, the local trivialization can be given by both φ_α and φ_β:

[4] Notice, however, that the fibre is also a general manifold, not necessarily a set of straight lines. This simply allows for a less confusing picture.

$$(U_\alpha \cap U_\beta) \times F \xleftarrow{\varphi_\alpha} \pi^{-1}(U_\alpha \cap U_\beta) \xrightarrow{\varphi_\beta} (U_\alpha \cap U_\beta) \times F \tag{2.12}$$

Even though the image is the same for both these maps, the considered trivializations may not map exactly to the same values. In order to relate these, we defined transition functions from $(U_\alpha \cap U_\beta) \times F$ to $(U_\alpha \cap U_\beta) \times F$ as follows:

$$\begin{aligned} \lambda_{\alpha\beta} : (U_\alpha \cap U_\beta) \times F &\longrightarrow (U_\alpha \cap U_\beta) \times F \\ (p, f_p) &\longmapsto \lambda_{\alpha\beta}(p, f_p) = (p, f_p \cdot g(p)) \end{aligned} \tag{2.13}$$

where $\lambda_{\alpha\beta} \equiv \varphi_\beta \circ \varphi_\alpha^{-1} = \varphi_\alpha \circ \varphi_\beta^{-1}$ are the transition functions, and $g(p) \in G$ is an element of the strucutral group. Equation (2.13) can be interpreted as follows: the coordinates in the base space are the same, since it's the same point, but the fibre element associated to it might be different under two different local trivializations. These can be compared with one another under the action of transition functions, which are elements of the structural group. If we let $f_{p,\alpha}$ and $f_{p,\beta}$ denote the fibre elements associated with the restrictions $(\varphi_\alpha \circ \pi^{-1})(U_\alpha \cap U_\beta)$ and $(\varphi_\alpha \circ \pi^{-1})(U_\alpha \cap U_\beta)$, respectively, then we have:

$$f_{p,\beta} = f_{p,\alpha} \cdot g(p) \tag{2.14}$$

representing transitions functions via an element of the structure group, thus effectively "gluing" together different subsets on the overlap region (see Figure 2.5).

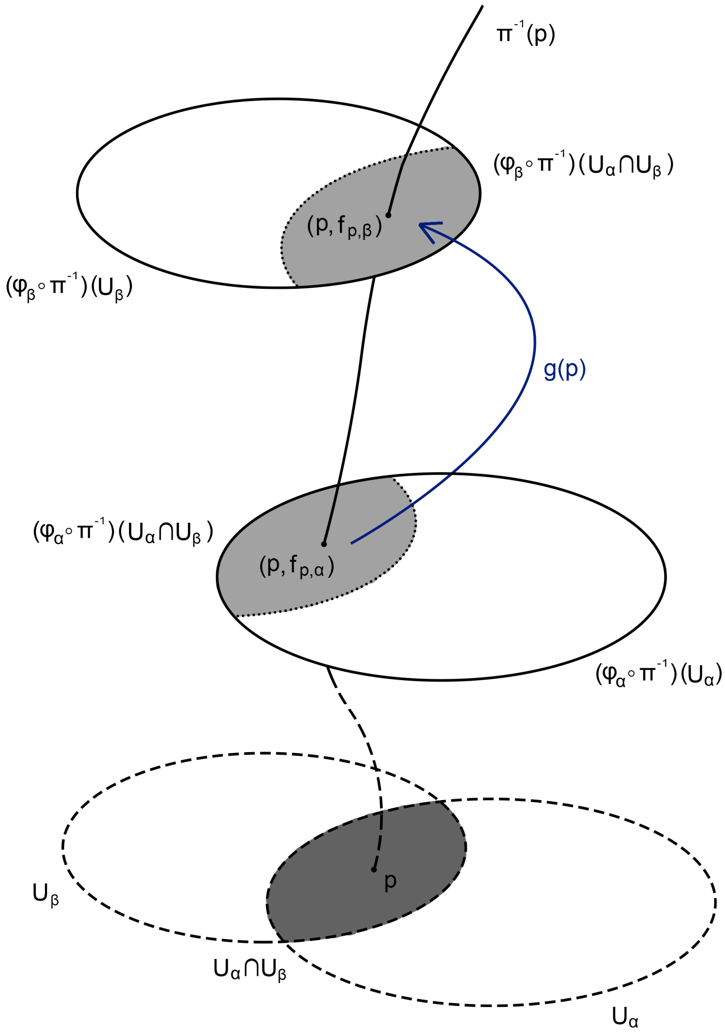
This construction is quite abstract, but it can be tackled by thinking of it as a way of formalizing the notion of extra degrees of freedom regarding some element in M. Finally, we define a local section of the total space as being a smooth map:

$$\begin{aligned} \sigma : M &\longrightarrow E \\ p &\longmapsto \sigma(p) = (p, f_p) \end{aligned} \tag{2.15}$$

where $f_p \in F_p$, associating with a point in the base space, a certain element of the fibre parametrized by that point (Marsh, 2019). Notice that the section, also called cross-section, "chooses" from the fibre associated to that point, a single element.

A special kind of fibre bundle is one where the fibre F is itself the structure group G, (consequently, a Lie group). This is called a G-principal fibre bundle, and it will be the focus of our discussion. Also, if the total space can be globally considered as a product $F \times G$, then we say that the fibre bundle is trivial.

Figure 2.5: Local representation of the total space E in an overlapping region and associated transition functions acting on fibre elements.



Source: Adapted from Marsh (2019).

2.4 Connections on the Fibre Bundle[5]

Consider the total space E. As was stated before, it can be locally approximated to the product of two manifolds: a subset of the base space M and the fibre F, which is also, for our purposes, a Lie group. Then, the projection is locally given by the following map:

$$\pi|_U : U \times F_U \longrightarrow U \tag{2.16}$$

where U is a coordinate neighbourhood of M around the point $p = \pi(\tilde{p})$, where $\tilde{p} \in E$,

[5] This subsection content is due mostly to Isham (1999), Way (2010) and Marsh (2019). Also, here we make use of vectors to describe the need for a connection and covariant derivation, but these concepts can be constructed much more generally. However, ease of understanding is lost for the sake of abstraction, which is not our intention. For this reason, we present a vector description, even though we won't use it, fully, later on.

and $\pi|_U$ is the restriction of the projection map to that subset. If we take the push-forward introduced in subsection 2.2.2 and apply it to this projection map, we obtain:

$$\pi_* : T_{\tilde{p}}E \simeq T_pM \times T_{\tilde{p}}F \longrightarrow T_pM \tag{2.17}$$

taking tangent vectors of the total space and mapping them on the tangent space of the base space at some point. Here, we can define the vertical subspace of the total space as the kernel of π_*:

$$V_{\tilde{p}}E = \ker(\pi_*) = \{ v \in T_{\tilde{p}}E \,|\, \pi_* v = 0 \} \tag{2.18}$$

that is, the space of tangent vectors to the total space which have no projection onto the tangent space of the base manifold. It is denoted as the vertical subspace, because it is the tangent space of the fibre, which is visualized as vertical, since its projection onto the base space is a single point. To be noticed that, even though the vertical and horizontal subspaces are complementary of one another, they are not orthogonal (in this abstract geometric object, there's no notion of angles to begin with).

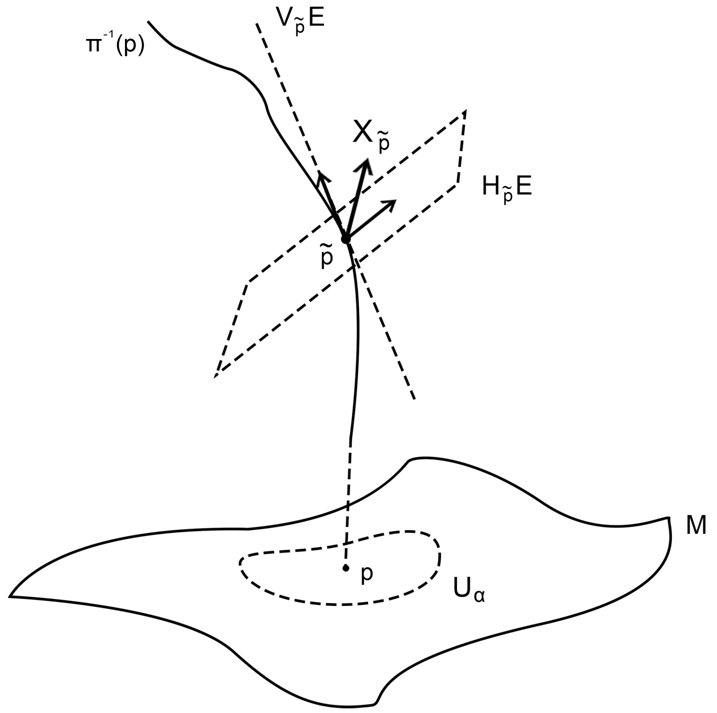
Hence, we can decompose the tangent space $T_{\tilde{p}}E$ at a point $\tilde{p}$ as the direct sum $\oplus$ of the vertical subspace $V_{\tilde{p}}E$ and a complementary space, of dimension $\dim(T_{\tilde{p}}E) - \dim(V_{\tilde{p}}E)$, called horizontal subspace $H_{\tilde{p}}E$ (see Figure 2.6). Notice that the horizontal subspace can also be defined as $\pi_*(H_{\tilde{p}}E) = T_{\pi(\tilde{p})}M$. Therefore, we have:

$$T_{\tilde{p}}E \simeq H_{\tilde{p}}E \oplus V_{\tilde{p}}E \tag{2.19}$$

The vector subspace at a point in the fibre is unique, but there are plenty of possible horizontal subspaces to be chosen. It is this attribution of a horizontal subspace for every point in the fibre that constitutes a connection, the necessary tool for comparing vectors between fibres. More formally, a connection in a principal bundle is a smooth assignment $\tilde{p} \to H_{\tilde{p}}E$ such that:

- $T_{\tilde{p}}E \simeq H_{\tilde{p}}E \oplus V_{\tilde{p}}E$, it is the vertical subspace complementary;
- $R_{g*}(H_{\tilde{p}}E) = H_{R_g(\tilde{p})}E = H_{\tilde{p}\cdot g}E$, "translations" of the horizontal subspace via the push-forward of the right action is equal to the horizontal subspace at the new point.

Figure 2.6: Decomposition of the tangent space of E at a point $\tilde{p}$



Source: Adapted from Way (2010).

From this definition, it can be proven that a connection 1-form [6] with values in the Lie algebra can be defined, here denoted as $\omega : T_{\tilde{p}}E \to \mathscr{G}$ [7]. Essentially, it maps the vertical component of a vector tangent in $T_{\tilde{p}}E$ to the tangent space of the fibre, which is a Lie algebra (see subsection 2.2.2). In order for this map to be a connection 1-form, two conditions must be met, which are similar to the ones above, but reformulated in terms of differential forms instead of vector subspaces.

[6] 1-forms are fundamental concepts in differential geometry, which, due to their complexity, are not present in our account of the formalism. However, a small explanation can still be provided: a 1- form is an (smooth) assignment of a cotangent vector ω_p to each point $p \in M$; in turn, a cotangent vector is any linear map $T_pM \to \mathbb{R}$, that is, which maps vectors into scalars; the set of all possible linear maps of this sort create the cotangent space at $p \in M$, T_p^*M, which is the dual space to T_pM. In practice, a 1-form in local coordinates can be written as a linear combination of basis cotangent vectors, denoted as dx^i, which act on the basis vectors ∂_i as follows: $dx^i(\partial_j) = \delta^i_j$, where δ^i_j is the Kronecker delta. The generalized version of a 1-form is called a k-form, or differential form, which is a multi-linear map $T_pM \times \cdots \times T_pM \to \mathbb{R}$. A 0-form is simply a function.

[7] There exists in fact an isomorphism between the Lie algebra and the vertical subspace, even though it is not shown here. However, it can be understood intuitively, since the vector subspace is defined as the space of vectors with no projection onto the base manifold, exactly because they belong to the tangent space of the fibre, a Lie algebra.

2.5 Parallel Transport and Covariant Derivative[8]

The construction of a connection is quite abstract, especially for our future purposes, but it allows for the following intuitive explanation of how two elements of different fibres may transform from one to the other. Consider an infinitesimal path $\tilde{\gamma}$ in the total space E, parametrized by $t \in [0,T]$, connecting two points $\tilde{p} \equiv \tilde{\gamma}(0)$ and $\tilde{q} \equiv \tilde{\gamma}(T)$. We can visualize point $\tilde{p}$ "following" a tangent vector to the curve $\dot{\tilde{\gamma}}$, for all t, until it reaches point $\tilde{q}$.

Projecting each point of the curve to the base space, we obtain $\pi(\tilde{\gamma}(t)) = \gamma(t)$. Now, there are two possible paths in the total space: a section of E along the curve γ such that $\sigma(\gamma(t)) = \tilde{\gamma}(t)$; and a horizontal lift of γ, which is a curve $\gamma_{||}(t)$ in the total space E such that $\pi(\gamma_{||}(t)) = \gamma(t)$ and $\dot{\gamma}_{||}$ lies entirely in the horizontal subspace determined by the connection $H_{\tilde{\gamma}(t)}E$ (see Figure 2.7). There's no *a priori* reason to assume they'll be equal.

So, the fibre element $\tilde{p}$ can be "guided" by tangent vectors to two versions of the same curve. Clearly, the endpoint will be different for both instances (see Figure 2.7). For the purpose of further highlighting the notion that the path $\tilde{\gamma}(t)$ is equal to some section of the path $\gamma(t)$ in the total space, defined for all points belonging to such a curve, we'll make use of $\sigma(p)$ to denote the fibre element $\tilde{p}$. Along the curve $\gamma_{||}(t)$, we know that its tangent vectors lie entirely in the horizontal subspace, determined by the chosen connection. Here we begin to notice the path-dependency of the connection, since a different path would result in different tangent vectors, which in turn would lie entirely in different horizontal subspaces. So, $\sigma(p)$ "follows" $\gamma_{||}$ until it reaches the endpoint $\gamma_{||}(T)$, always fulfilling the condition of $\dot{\gamma}_{||}(t)$ lying completely in the horizontal subspace. On the other hand, along the curve $\gamma_{||}(t)$, the element $\sigma(p)$ is "taken" to $\sigma(q)$. The horizontal subspace at this point is determined by the translation along the fibre, as previously stated. Thus, the tangent vector $\dot{\tilde{\gamma}}(t)$ will have a decomposition onto a vertical and horizontal subspace. We can visualize the vertical component as responsible for the upward "push", resulting in the displacement of the curve, and subsequent fibre element. The vertical component of $\dot{\tilde{\gamma}}$ is mapped onto the Lie algebra via the connection 1-form ω, already introduced. It can be proven that there exists a map which relates the Lie algebra with the Lie group itself (Isham, 1999).

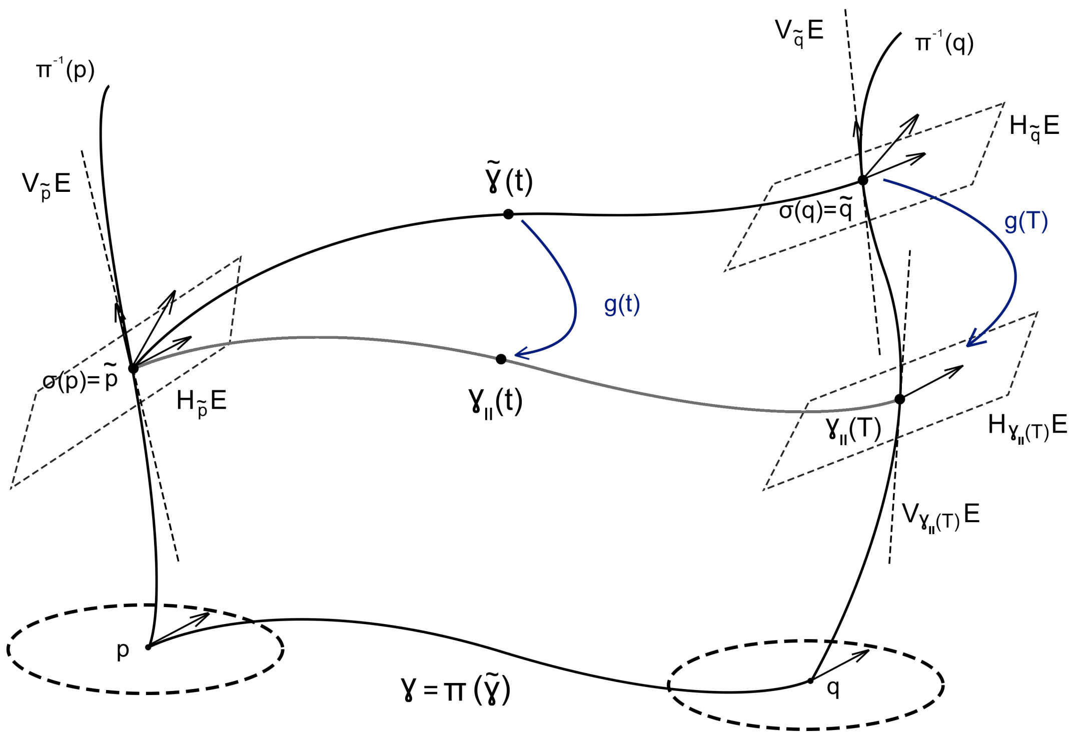
Furthermore, it can also be proved (Isham, 1999) that there exists an element g of the Lie group for each value of the parameter t, relating both paths $\tilde{\gamma}(t)$ and $\gamma_{||}(t)$:

$$\gamma_{||}(t) = \tilde{\gamma}(t)g(t) \tag{2.20}$$

[8] Both this and next sections were inspired mostly by Isham (1999), Way (2019) and Marsh (2019).

which allows for comparison between $\tilde{\gamma}(T) \equiv \sigma(q)$ and $\gamma_{\parallel}(T)$, given that they both belong to the same fibre $\pi^{-1}(q)$. It is this operator which defines the parallel transport of a fibre element across a path in the total space.

Figure 2.7: Two "versions" of the same path considered in the total space.



Source: Adapted from Marsh (2019).

By differentiating (2.20), and evaluating it in local coordinates, after some mathematical manipulations, an expression for $g(t)$ can be found (Isham, 1999):

$$g(t) = \mathrm{P}\exp\left\{-\int_0^t \left(\sum_{\mu} A_{\mu}(\gamma(s))\dot{\gamma}^{\mu}(s)\right) ds\right\} g(0) \tag{2.21}$$

The operator Pexp denotes a path-ordered exponential and $A_{\mu}(\gamma(s))$ is the μ component of the connection 1-form ω evaluated in local coordinates in the base space, parametrized by $s \in [0,t]$. If we assume that $\gamma_{\parallel}(0) = \tilde{\gamma}(0) = \sigma(p)$, then $g(0) = e$.

A path-ordered exponential is one where the order of operator's application matters:

$$A_{\mu}(\gamma(s))A_{\nu}(\gamma(s)) \neq A_{\nu}(\gamma(s))A_{\mu}(\gamma(s)) \tag{2.22}$$

Since we are only interested in Abelian groups, the above expression simplifies considerably to the usual exponential of an integral:

$$g(\gamma) = \exp\left\{-\int_{\gamma} A(\gamma)\dot{\gamma}^{\mu}(s)\right\} \tag{2.23}$$

where $A(\gamma) = \sum_\mu A_\mu(\gamma(t))d\gamma^{\mu}(t)$. This motivates the introduction of the following parallel transport map:

$$\begin{aligned} U_\gamma : \pi^{-1}(p) &\longrightarrow \pi^{-1}(q) \\ \sigma(p) &\longmapsto U_\gamma(\sigma(p)) = \sigma(q)\cdot g(\gamma) \end{aligned} \tag{2.24}$$

It essentially states that moving from fibre to fibre is achieved by "moving" along the path $\tilde{\gamma}(t)$ and applying $g(\gamma)$ at each t, to get to the path $\tilde{\gamma}(t)$. Or, formulated differently, moving along the path $\gamma(t)$ in the base space induces additional variations, only "visible" when one considers the total space, which are encoded in the changes along the fibre via $g(t)$. It's this operator that we'll effectively use in later sections. It abides by two conditions (Ilinski, 2001):

1. $g(\gamma_1\gamma_2) = g(\gamma_1)g(\gamma_2)$, applying the connection associated to the path $\gamma_1\gamma_2$ is the same as applying each connection individually, by order;
2. $g(\gamma^{-1}) = g(\gamma)^{-1}$, inverting the path is the same as applying the inverse.

Finally, the way the vectors themselves change along the curve can be captured by the notion of covariant differentiation. Essentially, it consists of differentiating them as usual, but with the addition of an adjustment, due to the changes in the associated fibres. It can be proven that the covariant derivative is as follows:

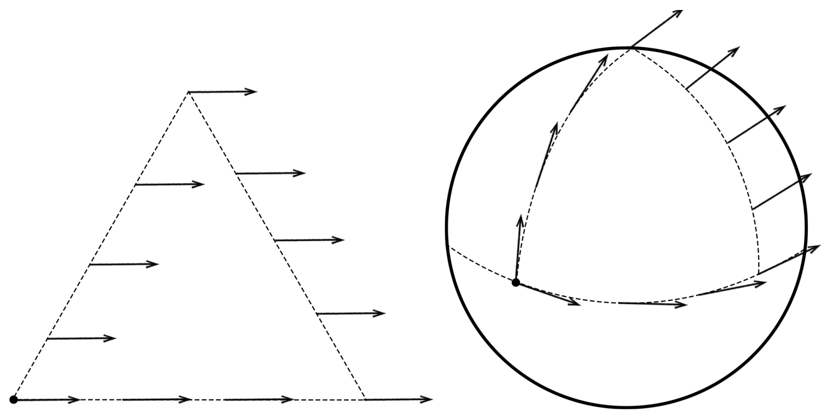
$$\nabla_\gamma = d + A(\gamma) \tag{2.25}$$

where d is the usual exterior derivative in local coordinates. The subscript γ emphasizes the observed path dependency.

2.6 Curvature

Curvature is an important concept in differential geometry for many reasons, one being the fact that it is a completely intrinsic quality of any given object, and as such, it must be independent of the coordinates used, as already stated in the beginning of this chapter. It is however, a complicated concept to be introduced formally, especially in the mathematical framework of differential geometry. For a completely heuristic approach, we begin by providing a classical example, which exposes quite nicely the notion of curvature, with the usage of parallel transport.

Figure 2.8: Parallel transport of a vector along (Left) the sides of an equilateral triangle in Euclidean space. (Right) and on the surface of a sphere.



Source: Adapted from Ilinski (2001).

Consider a sphere, embedded in $\mathbb{R}^3$. On its surface, we construct a path connecting three points, one on a pole and the other two in the "equator", all equidistant from one another, as pictured in Figure 2.8. If we take a tangent vector to the curve at some point (denoted in Figure 2.8 as a black dot), and parallel transport it along this closed path, always making sure that the vector belongs in its entirety to the horizontal subspace already defined via the chosen connection [9], we conclude that the final vector is not the same as that with which we've started. If we compare this with an equivalent closed-path, but in a flat surface (an equilateral triangle), we see that the same does not happen, and starting and ending vectors are the same. This change along a closed curve implies curvature [10]. It's with this notion of parallel transport along loops that we introduce curvature for the case of a principal fibre bundle.

Consider Figure 2.9. Here, we have a closed path connecting three points p, h and q in the base space, resultant from the projection of the associated paths in the total space. In the total space, we also have their respective horizontal lifts. Using parallel transport, we can relate the local sections $\sigma(p)$, $\sigma(h)$ and $\sigma(q)$ along each individual path as follows:

$$\begin{cases} U_{\gamma_1}(\sigma(p)) = \sigma(h) \cdot g(\gamma_1) = \sigma(h) \cdot \exp\left\{-\int_{\gamma_1} A(\gamma_1)\right\} \\ U_{\gamma_2}(\sigma(h)) = \sigma(q) \cdot g(\gamma_2) = \sigma(q) \cdot \exp\left\{-\int_{\gamma_2} A(\gamma_2)\right\} \\ U_{\gamma_3}(\sigma(q)) = \sigma'(p) \cdot g(\gamma_3) = \sigma'(p) \cdot \exp\left\{-\int_{\gamma_3} A(\gamma_3)\right\} \end{cases} \tag{2.26}$$

[9] This connection is usually known as a Levi-Civita connection. For our purposes, we may think of it as a connection which allows for a parallel transport that preserves angles (hence, the difference between the initial and final vectors in Figure 2.8).

[10] In fact, the Ambrose-Singer theorem provides a relation between curvature and holonomy, which is basically the group of all parallel transport operators (to be introduced later) acting on all possible loops (Kobayashi & Nomizu, 1963).

where $\sigma'(p)$ is an element of $\pi^{-1}(p)$, obtained by "following" $\tilde{\gamma}_3$. Notice that this is not necessarily equal to $\sigma(p)$. Also, we can parallel transport the fibre element $\sigma(p)$ along the loop γ, constituted by γ_1, γ_2 and γ_3, using the parallel transport operator $U_\gamma = U_{\gamma_3} \circ U_{\gamma_2} \circ U_{\gamma_1}$ previously introduced:

$$\begin{aligned} U_\gamma(\sigma(p)) &= \sigma'(p) \cdot g(\gamma) \\ &= \sigma'(p) \cdot \exp\left\{ -\oint_\gamma A(\gamma) \right\} \end{aligned} \tag{2.27}$$

where $g(\gamma) = g(\gamma_3) \cdot g(\gamma_2) \cdot g(\gamma_1)$ and $\oint_\gamma A(\gamma) = \oint_{\gamma_1} A(\gamma_1) + \oint_{\gamma_2} A(\gamma_2) + \oint_{\gamma_3} A(\gamma_3)$. By parallel transporting the fibre element $\sigma(p)$ along the horizontal lifted loop, under the operator U_γ, always taking in consideration the necessary adjustments, the mapping $\sigma(p) \mapsto \sigma'(p) \cdot g(\gamma)$ shows that the elements in the beginning and end aren't necessarily equal (see Figure 2.9), thus revealing the existence of curvature in the fibre bundle, similarly to the situation depicted in Figure 2.8.

Two conditions for the existence of curvature can be derived from equation (2.27). If $g(\gamma)$ is different than one, then there exists curvature. Clearly this can be generalized for a situation with more paths which, all together, give a closed loop. Hence, we can define the curvature $F(\gamma)$ with the following expression:

$$F(\gamma) = \left(\prod_{k=0}^{n} g(\gamma_{n-k}) \right) - 1 = g(\gamma) - 1 \tag{2.28}$$

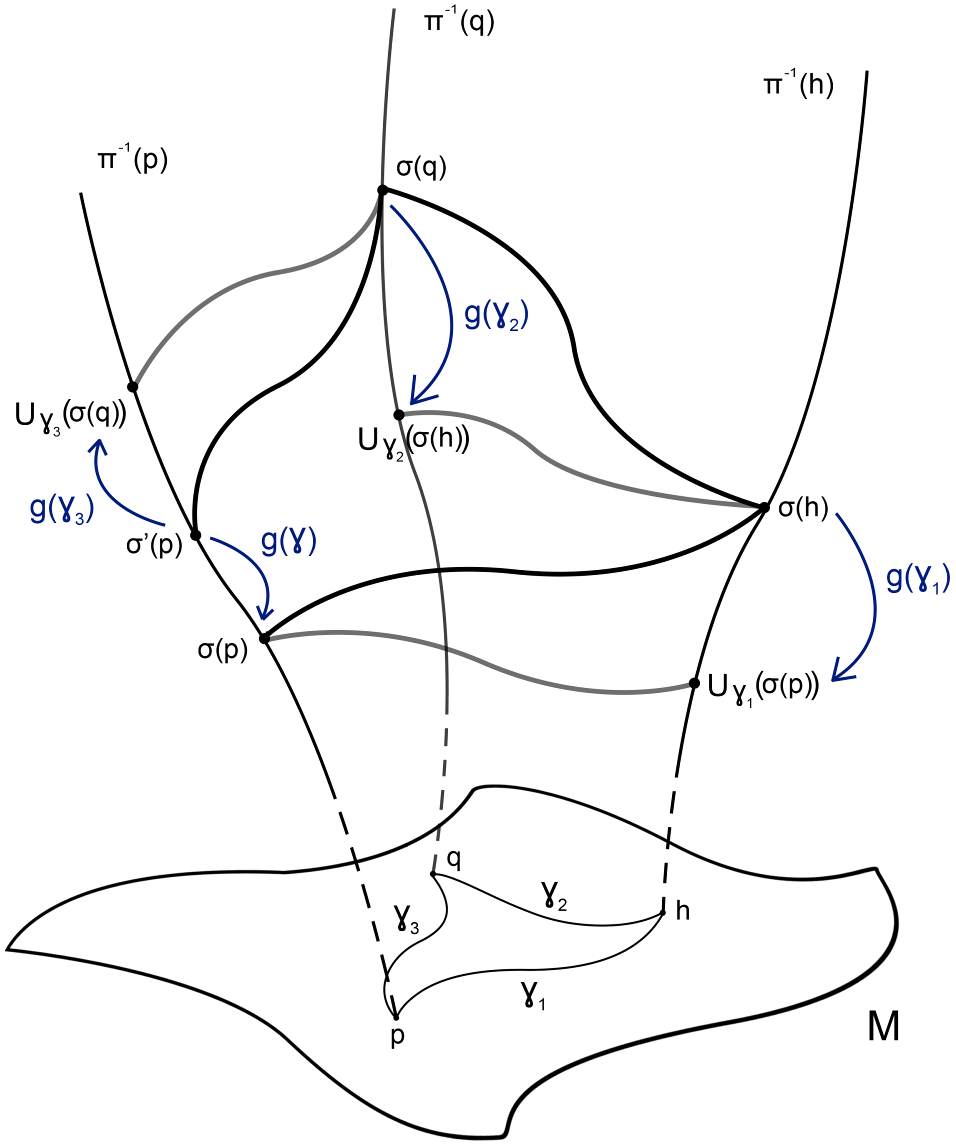
such that $\gamma = \bigcup_{k=0}^{n} \gamma_{n-k}$ is a loop. Also, if $-\oint_\gamma A(\gamma)$ is different than zero, there exists curvature. Applying the Stokes theorem [11], we conclude that the latter condition is equivalent to stating that if $\int_\Omega dA(\gamma)$ is different than zero, there exists curvature, where Ω is the area delimited by the loop. Hence, we can also define curvature with the following expression:

$$F(\gamma) = dA(\gamma) \tag{2.29}$$

For both cases, when $F(\gamma) \neq 0$, the fibre bundle is curved. The concept of curvature is mathematically expressed via the notion of a tensor, which can be briefly described as the multilinear map $T_pM \times \cdots \times T_pM \times T_p^*M \times \cdots \times T_p^*M \longrightarrow \mathbb{R}$. In particular, the curvature tensor is one whose "inputs" are zero vectors and two covectors. For this reason, it is also denoted as the curvature 2-form (see footnote 6).

[11] Notice that this does not work for the non-Abelian case.

Figure 2.9: Loop in the base space M and in the total space E.



Source: Adapted from Way (2010) and Marsh (2019).

2.7 Gauge Theory

With the mathematical formalism presented in this chapter, the tools to understand gauge theories have been acquired. What is presented below was based on the exposition encountered in François et al. (2014) and Marsh (2019).

A gauge theory can be loosely defined as any theory describing a system which can be transformed locally under some transformation (a gauge transformation), such that some quantities remain invariant, that is, exhibit gauge-invariance. Thus, we say that there exists gauge symmetry. It turns out that, considering physical systems, it's exactly these gauge-invariant quantities which are used to describe the system dynamics. An intuitive reason for this is the notion that real phenomena shouldn't be dependent on such arbitrary transformations.

Utilizing the mathematics exposed in the previous sections, gauge transformations are mediated through elements of the gauge group, which is the structural group acting on the fibres. Let $x \in M$ be an element of the manifold M. Let us define a gauge transformation as the following map:

$$\begin{aligned} \pi^{-1}(x) &\longrightarrow \pi^{-1}(x) \\ \sigma(x) &\longmapsto \sigma(x) \cdot q(x) \end{aligned} \tag{2.30}$$

where $q(x) \in G$, acting on fibre elements. Its dependence on x highlights the locality aspect. As we've seen, a connection on the fibre bundle can be understood as a way of comparing elements of different fibres. One of this formalism's great advantages is the non-reliance on coordinates, not in the sense that we don't express mathematical ideas using them, but that changes in coordinates are accompanied with changes in the mathematical ideas themselves. Hence, we want our connection to express the same information as it did before a gauge transformation. For that reason, it must follow a specific rule of transformation along with the transformation of the fibre elements.

Let the fibre elements $\sigma(x)$ and $\sigma(y)$ belong to fibres $\pi^{-1}(x)$ and $\pi^{-1}(y)$, respectively. Let them be connected (under parallel transport) as $U_\gamma(\sigma(x)) = \sigma(y) \cdot g(\gamma)$. If $\sigma(y)$ is subjected to a gauge-transform as $\sigma'(y) = \sigma(y) \cdot q(y)$, both $\sigma(x)$ and the connection must change in such a way that, under parallel transport, the endpoint is $\sigma'(y)$. Since the element $\sigma(x)$ changes to $\sigma'(x) = \sigma(x) \cdot q(x)$, the connection must change necessarily to $g'(\gamma) = q^{-1}(y) g(\gamma) q(x)$. Regarding its equivalent construction in terms of 1-forms, one can prove that each connection component must change as (Isham, 1999):

$$A_\mu \longrightarrow A'_\mu = A_\mu + q^{-1}(x) \frac{\partial q(x)}{\partial x^\mu} \tag{2.31}$$

We can immediately see that the connection elements do vary when one applies a gauge-transformation – the connection is not gauge-invariant. In other words, it depends on the coordinate system at use. Let us, however, consider a component of the curvature tensor (2.29). After some mathematical manipulations, one arrives at the following conclusion:

$$F_{\mu\nu} \longrightarrow F'_{\mu\nu} = q(x) F_{\mu\nu}\, q^{-1}(x) = F_{\mu\nu} \tag{2.32}$$

that is, the curvature is exactly a gauge-invariant quantity [12]. Therefore, it is independent of conventions, and it must be essential for the understanding of the dynamics underneath.

[12] Here we point out that, for the more general case of a non-Abelian group, the connection 1-form

Testimony for the usefulness of this formalism can be found all throughout physics' literature. A great deal of important fundamental theories have arisen from the notion of gauge symmetry. In fact, physics cements itself, at least partly, on the notion of symmetry. Important examples of this are the theory of electromagnetism, quantum mechanics and others (Ilinski, 2001).

components change in a different way, and the curvature is not gauge-invariant. However, quantities known as Wilson loops, which is exactly the trace of the parallel transport of fibre elements along loops, are gauge-invariant for any case (Ilinski, 2001).

3 METHODOLOGY

An important component of any scientific research is the literature review, whose objective is the synthesis of prior literature, to which the research pertains. In a typical literature review, the researcher chooses the scientific literature in such a way that provides a solid enough framework on which the new work is constructed. It is not required to detail all pre-existing scientific knowledge regarding the specific research topic at hand: all that is necessary is a setup for the actual work done by the researcher. This process of literature selection depends on the choices of the authors and it is therefore markedly subjective. However, when the research goal is exactly an account of all pre-existing knowledge and research done regarding some field, an objective literature review proves itself more suitable.

The need for a systematized method of conducting literature reviews first appeared in medical science, suggested by Cochrane (1999), providing a way of managing the ever-increasing, and sometimes, even contradictory, body of work outputted (Tranfield et al., 2003). As Tranfield et al. (2003, p.209) put it, "Systematic reviews differ from traditional narrative reviews by adopting a replicable, scientific and transparent process...", allowing for an objective account, minimizing the researcher's biases towards specific articles, of all prior knowledge regarding the chosen subject, in such a way as to be reproducible by any other researcher.

It would seem at first sight that one could simply apply the methodology put forward in the medical sciences to financial economics. However, we would not be taking into consideration the ontological and epistemological assumptions that differ between both disciplines (Durach et al., 2017). For this reason, we will follow instead the Tranfield et al. (2003) approach, where a "mapping" of this method from medical sciences to the management discipline was stipulated, much more closely aligned with finance and economics than the former.

The following three stages were undertaken when conducting the systematic literature review (Tranfield et al., 2003; Denyer & Tranfield, 2009): planning the review, where the research question was identified, discussed with a panel of specialists and a strategy to acquire the necessary data, delineated; conducting the review, where the strategy was employed, and the relevant data was filtered; finally, reporting and dissemination of findings, essentially discussion and conclusions arrived.

3.1 Planning the Review

The interest for the application of gauge theory to finance first appeared by the reading of the PhD thesis *The Index Number Problem: A Differential Geometric Approach*, by Malaney (1996), where differential geometry and gauge theory aspects are shown to be an appropriate language to describe economic reality.

This topic, and subsequent "search strategies" to be used, explained in the following sections, were submitted to extensive inspection and refinements by the review panel. The review panel was integrated by Dr. Luís Coelho, PhD in Management, Dr. Rúben Peixinho, PhD in Management, Dr. Nenad Manojlović, PhD in Theoretical Physics, and by Dr. Pedro Pintassilgo, PhD in Natural Resource Economics, and Dr. Cristina Viegas, PhD in Management, the two supervisors of this master's dissertation.

3.2 Conducting the Review

Given that we have no *a priori* notion of what might be interesting inside the formalism of gauge theory to apply to financial economics, we'll aim for complete generality, and since we are essentially applying concepts from one field to another, the search strings to be used will be only two, connected via the Boolean operator "AND": one based on gauge theoretical terms and another on financial terms.

3.2.1 Electronic Databases

Even though the followed protocol recommends the use of, not only articles and conference proceedings, but also working papers, we chose not to use the latter. The justification for this procedure lies in our aim to concentrate in peer-reviewed documents. This ensures the findings credibility. The same cannot be said about working papers, regardless of their potential validity and depth.

As a search engine, we utilized the Web of Science, which includes a large database of articles published in peer-review journals and conference proceedings.

3.2.2 Gauge Theoretical Keywords

Establishing keywords that best characterize the applications of gauge theory turns into a difficult endeavour, as it's associated to a great number of physical theories, and we are not interested in their application per se. Therefore, we only considered as keywords combinations of the word "Gauge" with "Theory", "Field", "Group", "Transformation", "Symmetry", "Invariant", "Fixing", "Independence", "Boson" and "Fermion".

3.2.3 Financial Keywords

Regarding the financial keywords, they were all extracted from the JEL Classification Codes Guide. Some modifications of these were, however, necessary in some cases. For example, the expressions “Financial Institution” or “Financial Service”, given that they share the word “Financial”, were shortened to a single one: “Finance”. This reasoning was applied every time that was possible to do so. Furthermore, any keyword which is ubiquitous in the English language was not considered for the search string. The explicit keywords used are presented in Appendix II.

3.2.4 Search String

A single search string was created. If we denote by “A” the combination of all gauge theoretical terms connected via the Boolean operator “OR”, and by “B” the same but for financial terms, then the final search string is simply “A AND B”.

The search was conducted by introducing the search string into the field “Topic” of the database, since it is the most general, comprising the title, abstract and the article keywords.

3.2.5 Paper Selection

After the search was completed, the next step was to determine which articles were relevant for our purposes. This must be done in such a way that preserves the systematic literature review’s objectiveness. The first stage was the application of exclusion principles to extract from our sample of papers, all which are in no way important. The second stage amounted to the application of inclusion principles to the already filtered sample, with the intention of constructing a final one where each element is of complete relevance. We now present the used principles.

3.2.6 Exclusion Principles

For exclusion criteria, the title and abstract of each paper were analysed. If keywords from both gauge theory and finance were present in the desired context, the paper would be included. Otherwise, excluded. Furthermore, only papers in the English language, and, as already mentioned, only articles and conference proceedings – papers submitted to peer-review – were considered.

3.2.7 Inclusion Principles

After the first filtering, we proceeded to the full-text reading of each paper, in order to assess its actual relevance. This final stage could be described as the most subjective. To minimize this tendency, we have directed our focus to the gauge theoretical formalism being applied, and did not include any application deemed as too superficial.

4 RESULTS

The search string previously mentioned was introduced in the database, by the 26^{th} of February, 2019. For the initial sample, 408 results were obtained. Based on our prior knowledge, fewer results were expected, given the relative novelty of the subject. An explanation for such a number is that some of the financial keywords, and the way that they were introduced in the database, appear quite pervasively in physics and mathematical literature, the fields where gauge theory first came to prominence. Alterations to these “problematic” keywords were attempted, but the fear of losing potentially relevant papers was a decisive factor to keep the string of keywords as it is.

We’ve proceeded to the exclusion criteria application. Despite the fact that the pool of results was vast, the great majority of the paper’s titles and abstracts were quite transparent regarding their relevance to our study. By applying the exclusion principles, we filtered the initial sample to 17 papers.

A quick look reveals an accumulation of papers around a specific subfield of finance: financial markets. In JEL’s notation, this subfield is characterized by the codes G1: G10, G11, G12, G13 and G17. Due to this reason, we’ve performed a more specific search, complemented with extra terms which were not present in JEL’s codes, but are related to them:

- Share Price, Share Pricing, Share Valuation
- Bond Price, Bond Pricing, Bond Valuation
- Option Valuation
- Put Option, Call Option
- Future Price, Future Pricing
- Forwards
- Swaps
- Spot Price
- Arbitrage
- Econophysics

This complementary search returned 12 results, all of them already in our previous list, with one exception, a conference paper. This exception was, however, relevant for our literature review. Hence, we’ve obtained 18 results.

Lastly, we applied the discussed inclusion principles, by carefully reading each paper. From our sample, four papers were not included, because of the fact that, even though gauge theoretical terms were present in the abstract, in many cases they seldom appeared anywhere else, and in the situations where they appeared, it was not in a context of interest to this literature review.

Table 4.1: Final sample of articles to be systematically reviewed.

Author(s) (Year)	Title
Sornette (1998)	Gauge Theory of Finance?
Young (1999)	Foreign Exchange Market as a Lattice Gauge Theory
Ilinski (2000)	Gauge Geometry of Financial Markets
Kholodnyi (2002a)	Valuation and Dynamic Replication of Contingent Claims in a General Market Environment based on the Beliefs-Preferences Gauge Symmetry
Kholodnyi (2002b)	Valuation and Dynamic Replication of Contingent Claims in the Framework of the Beliefs-Preferences Gauge Symmetry
Kholodnyi (2003)	Beliefs-Preferences Gauge Symmetry and Dynamic Replication of Contingent Claims in a General Market Environment
Morisawa (2009)	Toward a Geometric Formulation of Triangular Arbitrage – An Introduction to Gauge Theory of Arbitrage
Dupoyet, Fiebig & Musgrove (2010)	Gauge Invariant Lattice Quantum Field Theory: Implications for Statistical Properties of High Frequency Financial Markets
Sokolov, Kieu & Melatos (2010)	A Note on the Theory of Fast Money Flow Dynamics
Zhou & Xiao (2010)	An Application of Symmetry Approach to Finance: Gauge Symmetry in Finance
Dupoyet, Fiebig & Musgrove (2012)	Arbitrage-Free Self-Organizing Markets with GARCH Properties: Generating them in the Lab with a Lattice Model
Farinelli (2015)	Geometric Arbitrage Theory and Market Dynamics
Paolinelli & Arioli (2018)	A Path Integral based Model for Stocks and Order Dynamics
Paolinelli & Arioli (2019)	A Model for Stocks Dynamics based on a Non-Gaussian Path Integral

With this last criterion, we arrived to the final sample of results, 14 papers (see Table 4.1), to be discussed in the next section. A small synthesis of each paper is presented in the Appendix I, containing its objective, the nature of the gauge application and conclusions arrived.

Interestingly enough, all papers in the sample were published in physics related journals, begging the question of why hasn't this particular topic or application found a firm place in financial literature. We hypothesize that such an observation stems from the fact that a solid bridge between this particular physical mechanism and financial literature is lacking, despite the proven applicability and usefulness of various other physical theories and mathematical formalisms (Mantegna & Stanley, 2000), as stated in Chapter 1.

The sample spans across twenty-two years, with the largest number of papers concentrated in 2010. Furthermore, one can qualitatively describe these papers focuses as being aimed towards establishing solid theoretical grounds of financial markets dynamical behaviour.

From reading the papers, three different approaches were found, regarding the gauge theory application. One proposed by Kholodnyi in a working paper of 1995, here present in the 2002 article [13], where we have a beliefs-preferences gauge symmetry. The other approach suggested by Ilinski (1997) in the form of a working paper, subsequently criticized by Sornette (1998), and expanded by Young (1999) and Ilinski (2000), where we have a local *numéraire* gauge symmetry. The latter approach became the target of further investigation, by Ilinski (2001) himself, and also by other authors interested either in expanding these ideas, in theoretical terms – Morisawa (2009) and Zhou & Xiao (2010) – or by performing numerical simulations in order to understand to what extent is gauge invariance observed in financial markets – Dupoyet et al. (2010, 2012) and Paolinelli & Arioli (2018, 2019) –, or by providing important criticism – Sokolov et al. (2010).

The basis of the previous approach lead Farinelli (2015) to propose a completely geometric re-framing of stochastic finance, where gauge theory plays an important role, in a slightly different fashion. Here, a gauge symmetry regarding pairs of deflator and term structure (to be explained in Chapter 5, section 5.3) is explored, influenced by Smith & Speed (1998). Thus, Farinelli (2015) deviates from Ilinski's pioneer work, and therefore, it reserves a spot as a unique gauge theoretical approach to finance, different from the ones previously mentioned.

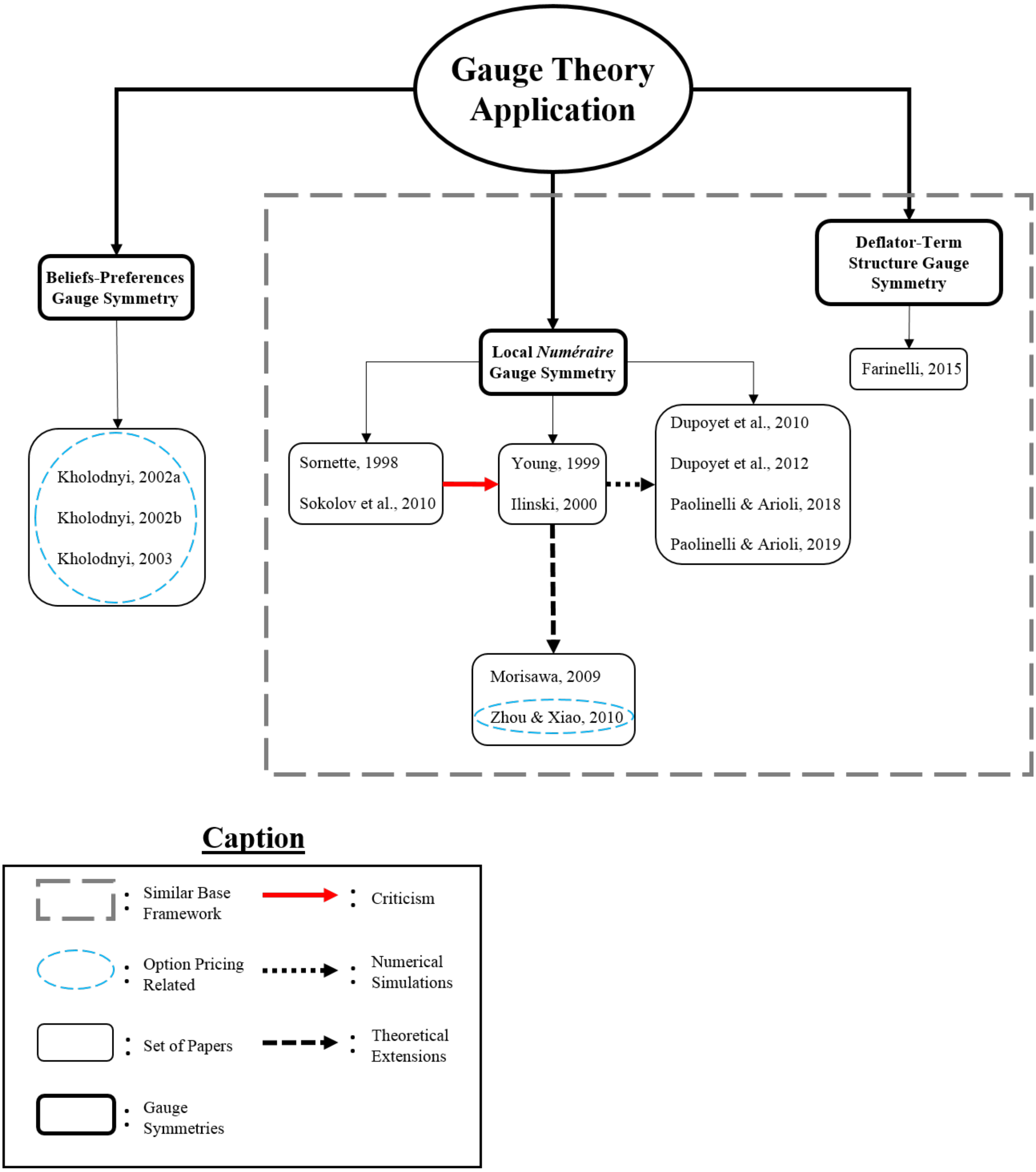
Last paragraph motivates the following differentiation, based on the type of gauge symmetry used: Kholodnyi (2002b) and Farinelli (2015) "approaches", denoted as

[13] In the list there are three papers by Kholodnyi, but all of them contain basically the same information, except the 2003 one, which is simply a shortened version. For this reason, a single article was considered: Kholodnyi (2002b).

“Beliefs-Preferences Gauge Symmetry” and “Deflator-Term Structure Gauge Symmetry” approach, respectively; all other articles pertain to a single approach, denoted from now on as “Local *Numéraire* Gauge Symmetry”, introduced by Ilinski (1997).

These observations are systematized in Figure 4.1, exposing as well the interconnectedness of the obtained articles: not only are the connections between articles pertaining to the “Local *Numéraire* Gauge Symmetry” approach illustrated, but also the notion that both “Local *Numéraire* Gauge Symmetry” and “Deflator-Term Structure Gauge Symmetry” approaches share a similar base framework; also, the “Beliefs-Preferences Gauge Symmetry” approach is represented as having little to nothing in common with the other two approaches, while maintaining a similarity with an article from the “Local *Numéraire* Gauge Symmetry” approach – Zhou & Xiao (2010) –, since these are the only ones applied to option pricing.

Figure 4.1: Diagram illustrating the set of pertinent papers and their applications of gauge theory to finance.



5 DISCUSSION

5.1 Beliefs-Preferences Gauge Symmetry

In this approach, Kholodnyi (2002b) focuses on European option pricing for general markets, whose price dynamics are not necessarily described by Markov processes. A market with a single underlying asset, whose price at time t is S_t, where transactions costs are non-existent and trading is always allowed, is considered.

The author begins by considering a European contingent claim starting at t and expiring at T whose payoff is $v_T(S_T) = v(T, S_T)$ [14].

Let the value of the contingent claim at time t be obtained via an evolution operator $V(t,T)$ as follows:

$$v_t = V(t,T)v_T \tag{5.1}$$

This operator completely characterizes the market environment. Let the utility of holding a unit of the European contingent claim be:

$$u'(t)v_t = F(t,T)(u'(T)v_T) \tag{5.2}$$

where $u'(t) = \partial u(t,S_t)/\partial v > 0$ is the marginal utility of consumption of one unit of European contingent claim and $F(t,T)(\cdot) = \int_0^\infty (\cdot)F(t,T,S_t,dS_T)$ is an operator denoting the transition probability of prices of the underlying asset. The expression above is interpreted as follows: the utility associated with a unit of the European claim at t is given by the utility of its expected payoff at T:

$$\begin{aligned} u'(t)v_t &= \mathbb{E}[u'(T)v_T] = \\ &= \int_0^\infty (u'(T)v_T)F(t,T,S_t,dS_T) = F(t,T)(u'(T)v_T) \end{aligned} \tag{5.3}$$

These operators describe each market participant, where $F(t,T)$ and $u'(t)$ represent their beliefs and preferences, respectively. An expression relating the market environment with its participants is easily derived, following from (5.1):

$$V(t,T) = (u'(t))^{-1}F(t,T)u'(T) \tag{5.4}$$

[14] A European contingent claim is a portfolio constituted of a long position on a European option with payoff $v_T^+(S_T) = max\{v_T(S_T),0\}$ and a short position on a European option with payoff $v_T^-(S_T) = -min\{v_T(S_T),0\}$. The total payoff is $v_T(S_T) = (v_T^+ - v_T^-)(S_T)$.

By the no-arbitrage argument, it is clear that v_t must be perceived equally amongst all market participants, and consequently, so is the evolution operator $V(t,T)$ equal for all participants. If this were not the case, and $V(t,T)$ depended on each participant's beliefs and preferences, the value of the contingent claim would be different among the participants, and these differences between values would lead to arbitrage opportunities, which would eventually return expression (5.4). In other words, in an equilibrium scenario, (5.4) must hold, regardless of participant's beliefs and preferences.

Hence, $(u'(t))^{-1}F(t,T)u'(T)$ must be the same for all participants, which implies a symmetry regarding changes in the beliefs and preferences of market participants. For a strictly positive function $g(t) = g(t,S_t)$, if beliefs and preferences transform as:

$$\begin{aligned} F(t,T) &\xrightarrow{g} g^{-1}(t)F(t,T)g(T) \\ u'(t) &\xrightarrow{g} g^{-1}(t)u'(t) \end{aligned} \tag{5.5}$$

then $V(t,T)$ remains invariant. Therefore, the market has a beliefs-preferences gauge symmetry. Let's assume that $V(t,T)$ and $F(t,T)$ are, when applied to v_t, solutions of the following differential equations:

$$\begin{cases} \frac{d}{dt}v(t) + L(t)v(t) = 0 \\ \\ \frac{d}{dt}v(t) + \ell(t)v(t) = 0 \end{cases} \quad , \; t < T, v(T) = v_T \tag{5.6}$$

where $L(t)$ and $\ell(t)$ are one-parameter operators which generate $V(t,T)$ and $F(t,T)$, respectively (essentially, infinitesimal counterparts). Equation (5.6) denotes general diffusion processes, where the operators aren't necessarily Markov operators (Kholodnyi, 2002b). Then the relation between market environment and its participants gains a new form:

$$L(t) = (u'(t))^{-1}\ell(t)u'(T) + \frac{\partial_t u'(t)}{u'(t)} \tag{5.7}$$

and so does the differential equation that models the European contingent claim's value evolution:

$$\begin{aligned} &\left(\frac{d}{dt} + \ell(t)\right) u'(t)v(t) = 0, \qquad t < T \\ &v(T) = v_T \end{aligned} \tag{5.8}$$

Under the gauge transformation[15]:

$$\begin{aligned} \ell(t) &\xrightarrow{g} g^{-1}(t)\ell(t)g(T) + \frac{\partial_t g(t)}{g(t)} \\ u'(t) &\xrightarrow{g} g^{-1}(t)u'(t) \end{aligned} \tag{5.9}$$

$L(t)$ and equation (5.8) remain invariant. However, if the gauge transformations acting on the beliefs and preferences of a participant are different from each other, $L(t)$ and equation (5.8) are not, in general, invariant. In fact, let the gauge-transformation acting on the pair belief-preference $(\ell(t), u'(t))$ be $(g(t), h(t))$. Then we have:

$$\begin{aligned} &\left(\frac{d}{dt} + \ell(t)\right) u'(t)v(t) = 0 \xrightarrow{(g,h)} \\ &\xrightarrow{(g,h)} \left(\frac{d}{dt} + q^{-1}(t)\ell(t)q(t) + \frac{\partial_t q(t)}{q(t)}\right) u'(t)v(t) = 0 \end{aligned} \tag{5.10}$$

where $q(t) = g(t)h^{-1}(t)$, which is also an element of the gauge group (see Chapter 2, subsection 2.2.2). Evidently, the equation loses invariance under independent gauge transformations. The goal becomes to model the value of a European contingent claim, independent of beliefs and preferences, individually.

With that in mind, Kholodnyi (2002b) first proposed approximating the belief operator through a series of differential operators with respect to the price of the underlying asset [16]:

$$\ell(t) \approx \ell\left(t, S, \frac{\partial}{\partial S}\right) = \sum_{m=1}^{M} \ell_m(t,S) \frac{\partial^m}{\partial S^m} \tag{5.11}$$

where $\ell_m(t,S)$ are the operator coefficients, obtained through the non-commutation relationship between $\ell(t,S)$ and S, which represents randomness.

Secondly, in order for equation (5.10) to be gauge-invariant, the derivatives themselves must be changed into their covariant counterparts. To achieve this, the geometry of the market fibre bundle, whose mathematics were introduced in Chapter 2, must be brought to the surface: the set of pairs "price" and "time" denote the base space, while the fibre is the set of all possible option' values obtained via the generator in (5.8), and as such,

[15] There's an important restriction on the kind of functions suitable to be gauge transforms. This restriction stems from the fact that the operator $F(t,T)$, being a transition probability meant to represent the participant's beliefs, acts on the identity $\mathbb{I}$ as $F(t,T)\,\mathbb{I} = \mathbb{I}$, which in turn implies that $l(t)\,\mathbb{I} = 0$ To ensure this, it is necessary that the elements of the gauge group are solutions of equation (5.6).

[16] This is known as the method of quasi-differential operators (Kholodnyi, 2002b).

can be identified with the generator itself, thus being the “target” of the structure group already introduced. Then, if the derivatives transform as follows:

$$\begin{aligned} \partial_t &\xrightarrow{q} \nabla_t = \partial_t + \frac{\partial_t q(t)}{q(t)} \\ \partial_S &\xrightarrow{q} \nabla_S = \partial_S + \frac{\partial_S q(t)}{q(t)} \end{aligned} \tag{5.12}$$

the evolution equation (5.10) acquires a covariant form:

$$(\nabla_t + \ell(t, S, \nabla_S))\, u'(t)v(t) = 0 \tag{5.13}$$

meaning that it remains invariant under gauge-transformations. These “corrections” to the derivatives are endowed with a very suggestive interpretation. One can easily understand that the gauge transformation is meant to represent possible changes to the beliefs and preferences which characterize a market participant. As we’ve seen, these changes shouldn’t affect the value of the contingent claim, and this is accomplished by balancing the way the contingent claim changes throughout time and due to changes of the underlying asset’s price. The way price changes is balanced by the relative change of beliefs and preferences with respect to price, and the way the value of the contingent claim varies through time is balanced by how the beliefs and preferences of the market participants change with time, thus reflecting the notion of participants influencing the price and value of market securities due to their behaviour.

Kholodnyi (2002b) proceeds to the calculation of this equation by utilizing a portfolio of other European contingent claims that replicates its evolution. Consider a portfolio of $N+1$ other European contingent claims on the same underlying asset, whose vaues at t are $v_\omega(t)$, for an index $\omega = 0, 1, ..., N$. Let the utility provided by the European contingent claim v_t evolve as follows:

$$\left[\frac{d}{dt} + \ell(t)\right] u'(t)v(t) = \sum_{\omega=0}^{N} \pi_v(t, \omega) \left[\frac{d}{dt} + \ell(t)\right] u'(t)v_\omega(t) \tag{5.14}$$

where $\pi_v(t, \omega)$ are the associated portfolio weights. In other words, the utility provided by the contingent claim evolves similarly to the utility of each of the constituent contingent claims, multiplied by some weight. If also:

$$v(t) = \sum_{\omega=0}^{N} \pi_v(t, \omega) v_\omega(t) \tag{5.15}$$

then the portfolio dynamically replicates the European contingent claim. Let the pure discount bond, indexed by $\omega = 0$, be in the set of replicating contingent claims. It can be shown that the above equation can be written in such a way that preferences do not appear:

$$\left[\frac{d}{dt}+\ell(t)\right]v(t)=\sum_{\omega=0}^{N}\pi_v(t,\omega)\left[\frac{d}{dt}+\ell(t)\right]v_\omega(t)\\ +\left(v-\sum_{\omega=1}^{N}\pi_v(t,\omega)v_\omega(t)\right)r(S,t) \tag{5.16}$$

where $r(S,t)$ is the bond's interest rate. For the above equation to be valid, the following condition must be met:

$$\left[[\ell(t),v],u'(t)\right]\mathbb{I}=\sum_{\omega=1}^{N}\pi_v(t,\omega)\left[[\ell(t),v_\omega],u'(t)\right]\mathbb{I} \tag{5.17}$$

which permits the derivation of the portfolio weights. If each contingent claim i, for where $i=0,1,...,N$, is itself a portfolio of j contingent claims, where $j=0,1,...,i$, with associated portfolio weights $\binom{i}{j}(-1)^{i-j}S_t^{i-j}$, then its value at t is:

$$P_i(t)=\sum_{j=0}^{i}\binom{i}{j}(-1)^{i-j}S_t^{i-j}\exp\left\{-\int_t^t\Omega_j(S,\tau)d\tau\right\}S_t^j \tag{5.18}$$

where $\Omega_j(S,\tau)$ can be interpreted as the term structure. Its relationship with $V(t,T)$ and $L(t)$, respectively, is as follows:

$$\Omega_j(S,t)S_t^j=\frac{d}{dt}V(t,T)\bigg|_{t=T}S_T^j=\\ =L(t)S_T^j \tag{5.19}$$

In other words, if the contingent claim i is:

- $i=0$, $j=0$: nothing but a portfolio of a single bond;
- $i=1$, $j=0,\ 1$: a portfolio of a bond and the underlying asset;
- $i=2$, $j=0,\ 1,\ 2$: a portfolio of a bond, the underlying asset and some "second order" combination of the two;
- ...

and so on. Using the method of quasi-differential operators, one obtains the following expression:

$$\left[\frac{\partial}{\partial t}-\sum_{i=1}^{\infty}\left(\frac{S^i}{i!}\left(\sum_{j=0}^{i}\binom{i}{j}(-1)^{i-j}\Omega_j(S,t)\right)\right)\frac{\partial^i}{\partial S^i}-r(S,T)\right]v=0 \tag{5.20}$$

which denotes the evolution of the value of a European contingent claim, provided that it is dynamically replicable by the portfolio considered above. It becomes elucidative to consider a couple of its first terms:

- For $i=1,\ j=0,\ 1$:

$$\left[\frac{\partial}{\partial t}+S(\Omega_0(S,t)-\Omega_1(S,t))\frac{\partial}{\partial S}-r(S,T)\right]v=0 \tag{5.21}$$

  where $\Omega_0(S,t)$ and $\Omega_1(S,t)$ are interpreted as the return on the bond $r(S,t)\equiv r$ and the dividends accrued by the asset $d(S,t)\equiv d$, respectively. This expression can be interpreted as a very crude approximation of how a European contingent claim evolves - it is equal to the dynamics of its underlying assets, a bond and a share;

- For $i=2,\ j=0,\ 1,\ 2$:

$$\left[\frac{\partial}{\partial t}+(r-d)S\frac{\partial}{\partial S}+\frac{1}{2}(2d-r-\Omega_2(S,t))S^2\frac{\partial^2}{\partial S^2}-r\right]v=0 \tag{5.22}$$

  which is the classical Black-Scholes equation, as long as the term $(2d-r-\Omega_2(S,t))$ is interpreted as volatility $\sigma^2(S,T)$.

Therefore, the evolution equation above can be seen as a generalization of the kind of diffusion processes which are pervasive throughout financial literature. Notice that this diffusion process is non-Markovian, due to the terms for which the index i is greater than two. Furthermore, changes to the participants beliefs and preferences of the kinds here exposed are potentially negated if one considers the covariant derivatives instead, which take into account the relative change of the gauge transformation element, with respect to changes in the price and time. If this gauge symmetry is broken, the author suggests that a specific kind of arbitrage opportunity might be present.

5.2 Local *Numéraire* Gauge Symmetry

We begin by providing an account of the basic framework of the local *numéraire* gauge symmetry approach. It follows from Young (1999) and (mostly) Ilinski (2000) [17], which are quite alike.

[17] Also from Ilinski (1997) and Ilinski (2001).

5.2.1 Basic Framework

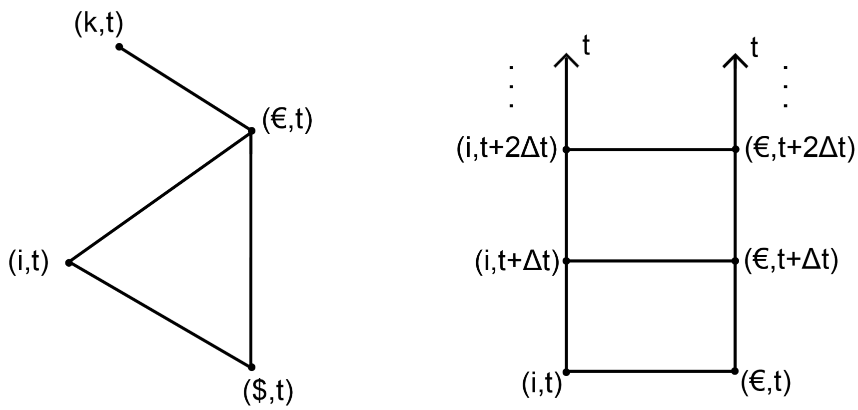
A general exchange market can be defined as a set of $N+1$ assets $I=\{i\,|\,i=0,\,1,\,2,...,\,N\}$, where the 0^{th} asset is interpreted as money in a bank account – cash. Time is introduced in a discrete fashion by assigning to each asset the set $T=\{t\,|\,t=...,-2,\,-1,\,0,\,1,\,2,...,\,N\}$. A space $M=I\times T=\{(i,t)\,|\,i\in I,\,t\in T\}$ of assets at each point in time is obtained, which can be visualized as a grid.

Assets that can be exchanged for one another must be "connected" in some way. To accomplish this, links between points in the grid must be introduced. These links must:

- Connect the same asset across time – an asset at present time can always be traded with the same asset in the future, which is equivalent to holding the asset across time;
- Connect all assets with the 0^{th} asset at all times – this is to say that every asset can be converted in money, which implicitly assumes well-functioning markets;
- Connect assets at some time that can be exchanged directly, e.g. currencies or derivatives at the spot time.

What was constructed here has the geometry of a "ladder", denoted in physics literature as lattice (see Figure 5.1).

Figure 5.1: (Left) Connected market of only four assets. (Right) Time evolution of two connected assets.



Source: Adapted from Ilinski (2001).

What is proven by Ilinski (2000) and serves as building block for the rest of this discussion, is the geometric nature of a general exchange market. This geometry is called a fibre bundle, as previously exposed in section 2.3.

Each element of M, an asset at some time, has to it associated a non-negative scalar, which reflects the amount of that asset owned by some economic agent. This scalar is an element of the fibre F, which, in our situation, is the infinite set $\mathbb{R}_+$. Together with M, a trivial fibre bundle $E = M \times F$ is constructed. Following the formalism introduced in Chapter 2, we have the projection map $\pi : E \longrightarrow M$ by $(x, f_x) \longmapsto x$, and cross-sections $x \longmapsto \sigma(x) = (x, f_x)$, which accounts for how much of asset i at time t, denoted as the pair $x = (i,t)$, does the investor hold, where $f_x \in F_x$ denotes that asset's amount.

Finally, the structural group G that acts on each element of the fibre via:

$$\begin{aligned} E \times G &\longrightarrow E \\ (p,g) &\longmapsto p \cdot g = (x, f_x \cdot g(x)) \end{aligned} \tag{5.23}$$

where p is an element of the total space E, and G is, once again, $\mathbb{R}_+$, the group of dilations. This is meant to reflect the changes that occur on the fibre elements, that is, the amount of each asset owned. Notice that, because the structure group is nothing but the set of scalars, the distinction between left and right action is meaningless here. Since $G = F$, our trivial fibre bundle is also called a principal fibre bundle (see section 2.3). Given a path between two points x_0 and x_n in the base space M, denoted as $\gamma = \{x_0, x_1, ..., x_n\}$, which is undertaken by some investor through trading (as long as there exists links connecting each asset-time point), it is also necessary to take into account the changes in the fibre elements. As was shown before in section 2.4, the choice of a connection is necessary to compare elements of different fibres. The comparison itself is realized by parallel transporting an element of the fibre along the curve until it reaches the endpoint. As we've seen, this is equivalent to saying that changes in the fibre element are solely attributed to the action of the group on it. Thus, we define the following parallel transport map U_γ, similar to that introduced in section 2.5, equation 2.24. However, because our fibre bundle is, in this case, discrete and Euclidean (a subset of $\mathbb{R}^N$), we will simplify future considerations by stating that to parallel transport an element of a fibre to another, one simply multiplies it by an element of the structural group pertaining to that specific path [18]. Therefore, we have:

$$\begin{aligned} U_\gamma : \pi^{-1}(x_0) &\longrightarrow \pi^{-1}(x_N) \\ \sigma(x_0) &\longmapsto \sigma(x_0) \cdot g(\gamma) = \sigma(x_N) \end{aligned} \tag{5.24}$$

[18] In subsection 2.5, the meaning of parallel transport is slightly different: it is an operator that takes an element $\sigma(x_0) \in \pi^{-1}(x_0)$ to $\pi^{-1}(x_N)$, and states that this new value is equal to a corrected $\sigma(x_N)$, that is, $U_\gamma(\sigma(x_0)) = \sigma(x_N) \cdot g(\gamma)$. Here, we mean to redefine this notion as follows: the parallel transport operator simply maps $\sigma(x_0) \in \pi^{-1}(x_0)$ directly to $\sigma(x_N) \in \pi^{-1}(x_N)$, by multiplying the former with an element of the structural group, dependent on the path, that is, $U_\gamma(\sigma(x_0)) = \sigma(x_N) = \sigma(x_0) \cdot g(\gamma)$. We do this to keep this section's notation similar to the one used in the references.

such that $U_\gamma = \prod_{i=0}^{N-1} U_{\{x_{N-i-1},\, x_{N-i}\}}$ and $g(\gamma) = \prod_{i=0}^{N-1} g(\{x_i,\, x_{i+1}\}) \in G$, where $\{x_{N-i-1},\, x_{N-i}\}$ and $\{x_i,\, x_{i+1}\}$ denote elementary portions of the path γ, connecting x_{N-i-1} to x_{N-i} and x_i to x_{i+1}, respectively, for all $i = 0,\ 1,\ \ldots,\ N-1$.

A financial interpretation is given (see Figure 5.2): connections in the financial fibre bundle are meant to represent prices and rates of return. Consider, for example, two elements of the base space M, (i,t) and $(0,t)$, that is, some share i and cash in some currency (we'll assume it's euros for simplicity) at time t, respectively. As it was shown, to each of these elements, there exists a fibre attached, equal to the non-negative portion of the set of real numbers (horizontal lines in Figure 5.2). Then, as elements of $E = M \times F$, we have $\sigma(i,t) = \left(i,t,f_{(i,t)}\right)$ and $\sigma(0,t) = \left(0,t,f_{(0,t)}\right)$, where $f_{(i,t)}$ and $f_{(0,t)}$ are the amounts of share and euros, respectively, that an economic agent may hold at time t. It is clear that the price in euros of this particular share is exactly the amount of euros that one gets for a single share. In differential geometric terms, it is the factor $g_{i,t}^{0,t} \in \mathbb{R}$ that permits the equality $\sigma(0,t) = \sigma(i,t) \cdot g_{i,t}^{0,t}$ to hold, for the path $\gamma = \{(i,t),(0,t)\}$ [19]. Hence, we notice that prices are exactly connections, and selling or buying [20] the asset is seen as a parallel transport.

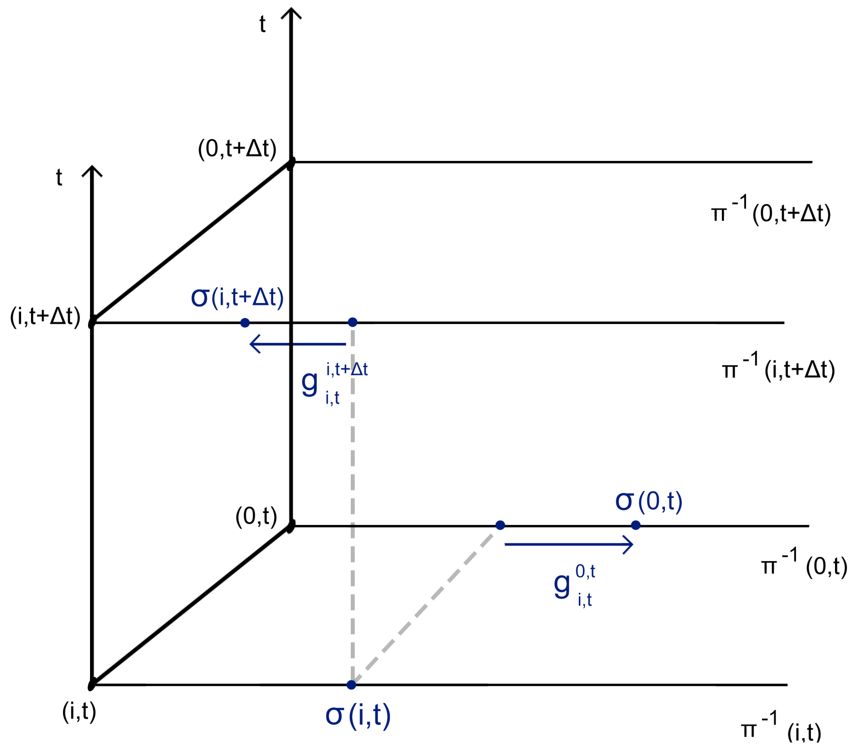
Also, exchange rates between currencies can be formalized in the same fashion: if we consider, instead of the base space element (i,t), the element $(0',t)$, where $0'$ denotes cash in a different currency, let's say dollars, then the factor $g_{0',t}^{0,t}$ is equal to the exchange rate between dollars and euros.

Another example can be given, now with respect to the evolution of assets across time. Let us consider the same asset, a generic share. If the investor decides to hold on to this share from the present time t to an immediate after period $t+\Delta t$, we know that the value of the share will change, formalized by the notion of rate of return. In differential geometric terms, we say that we are parallel transporting the element $\sigma(i,t)$ to $\sigma(i,t+\Delta t)$, under the connection $g_{i,t}^{i,t+\Delta t}$, by performing the operation $\sigma(i,t+\Delta t) = \sigma(i,t) \cdot g_{i,t}^{i,t+\Delta t}$, which is equivalent to continuous compounding the present value of the share. The opposite operation would be $\sigma(i,t) = \sigma(i,t+\Delta t) \cdot \left(g_{i,t}^{i,t+\Delta t}\right)^{-1}$, which is the share's future value discounted to the present – the Net Present Value.

[19] For ease of exposition, we set $g\left((i,t),(j,t)\right) \equiv g_{i,t}^{j,t}$ and $g\left((i,t),(i,t+\Delta t)\right) \equiv g_{i,t}^{i,t+\Delta t}$.

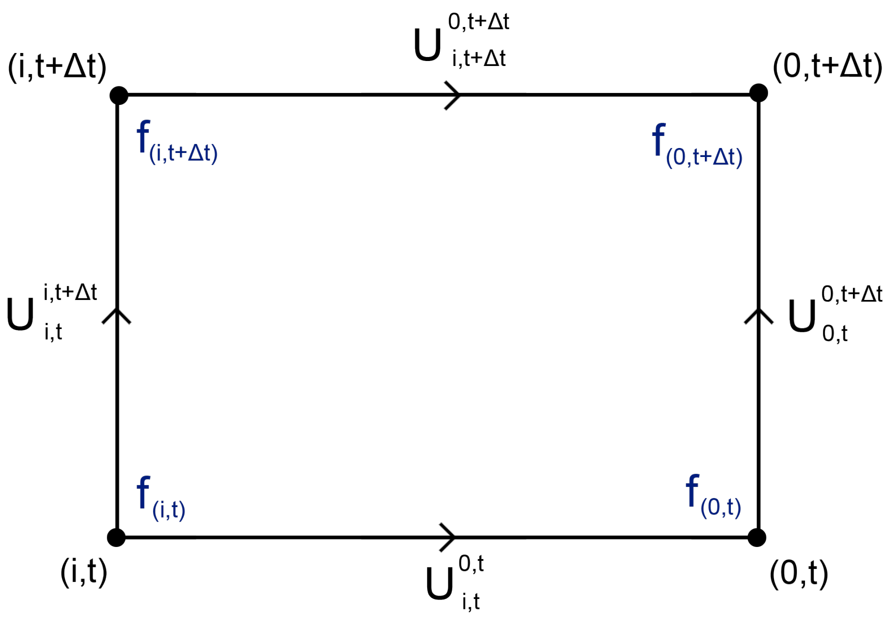
[20] Equivalent to the path $\gamma^{-1} = \{(0,t),(i,t)\}$, which, in an ideal scenario with no transaction costs or bid-ask spreads, is $g_{0,t}^{i,t} = \left(g_{i,t}^{0,t}\right)^{-1}$.

Figure 5.2: Fibre bundle geometry of a portion of the lattice - *plaquette*.



Returning to our framework, prices and/or exchange rates between assets (i,t) and (k,t) will be denoted as $U_{i,t}^{k,t}$, while interest rates for a particular asset (i,t), by $U_{i,t}^{i,t+\Delta t}$, for all assets in the considered market. Furthermore, from now on we are interested on the fibre element $f_{(i,t)}$, instead of $\sigma(i,t)$.

Figure 5.3: Elementary *plaquette*, with elements of the fibre in blue.



Source: Adapted from Ilinski (2001).

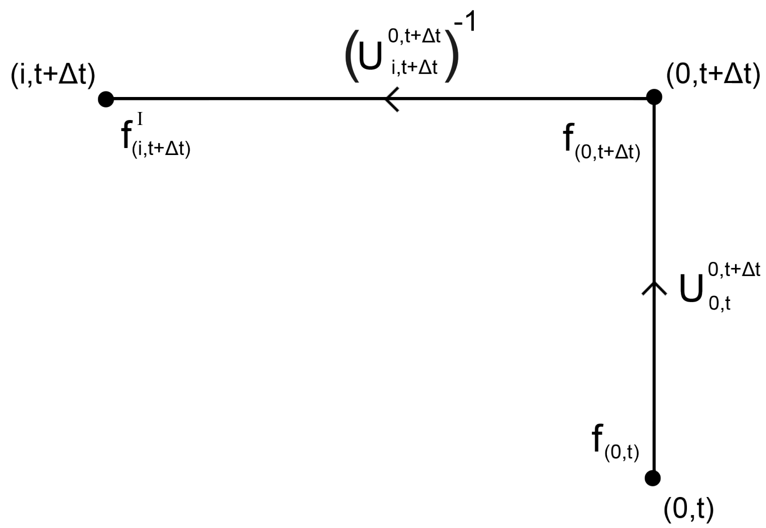
Consider two possible trading strategies in the same elementary loop on the lattice construction, to be denoted from now on as a *plaquette* (see Figure 5.3). For ease of exposition, let's say that one of the assets is cash in a bank account, denoted as $(0,t)$ for each time period t. Notice, however, that the more general case is equivalent. Furthermore, notice that the operators act on the right, even though this distinction is, as already stated, unimportant. We maintain it, however, in order to keep consistent with the mathematical literature.

As an investor with an amount $f_{(0,t)}$ of cash at time t, the two strategies are as follows:

I. Deposit cash in a bank account at some interest, formalized by the connection $U_{0,t}^{0,t+\Delta t}$, and buy asset i at $t+\Delta t$, at price $\left(U_{i,t+\Delta t}^{0,t+\Delta t}\right)^{-1}$. Then, at $t+\Delta t$ we have the following amount of asset i:

$$f_{(i,t+\Delta t)}^{I} = f_{(0,t)} U_{0,t}^{0,t+\Delta t} \left(U_{i,t+\Delta t}^{0,t+\Delta t}\right)^{-1} \tag{5.25}$$

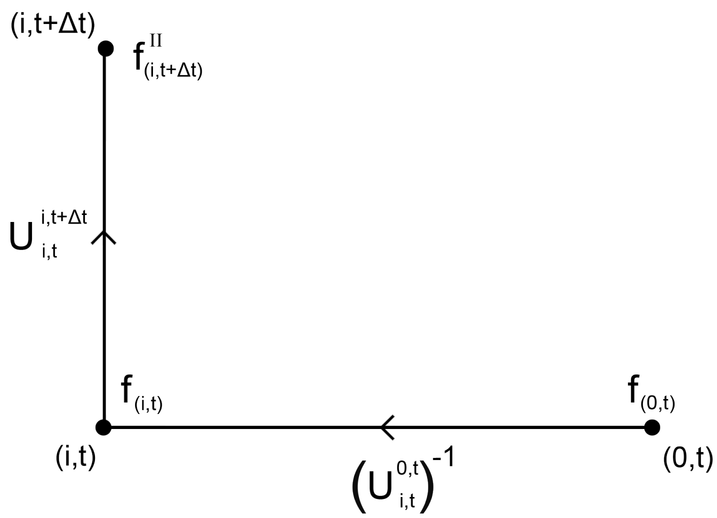
Figure 5.4: One possible arbitrage strategy, denoted as strategy I.



II. Buy asset i at price $\left(U_{i,t}^{0,t}\right)^{-1}$, and receive the return accrued by the asset at some rate, obtained via the connection $U_{i,t}^{i,t+\Delta t}$. Then, at $t+\Delta t$, we have the following amount of asset i:

$$f_{(i,t+\Delta t)}^{II} = f_{(0,t)} \left(U_{i,t}^{0,t}\right)^{-1} U_{i,t}^{i,t+\Delta t} \tag{5.26}$$

Figure 5.5: One possible arbitrage strategy, denoted as strategy II.



If $f^{I}_{(i,t+\Delta t)}$ and $f^{II}_{(i,t+\Delta t)}$ are different, it means that one of the strategies is clearly superior. Therefore, and assuming ideal markets, where borrowing is always possible, with no transaction costs, arbitrage opportunities can occur.

If $f^{I}_{(i,t+\Delta t)} > f^{II}_{(i,t+\Delta t)}$, a possible arbitrage strategy would be to borrow an amount $f_{(i,t)}$ of asset i and sell it for $f_{(i,t)}U^{0,t}_{i,t}$ (that is, to short sell asset i). Applying the first trading route, at $t+\Delta t$ one would gain $f_{(i,t)}U^{0,t}_{i,t}U^{0,t+\Delta t}_{0,t}\left(U^{0,t+\Delta t}_{i,t+\Delta t}\right)^{-1}$ amount of asset i, but because $f_{(i,t)}$ was borrowed at t, one owes $f_{(i,t)}U^{i,t+\Delta t}_{i,t}$ at $t+\Delta t$.

Therefore, this trading strategy profits:

$$\Pi^{(I)}(t+\Delta t) = f_{(i,t)}\left[U^{0,t}_{i,t}U^{0,t+\Delta t}_{i,t}\left(U^{0,t+\Delta t}_{i,t+\Delta t}\right)^{-1} - U^{i,t+\Delta t}_{i,t}\right] \tag{5.27}$$

which, discounted to time t, is equal to:

$$\Pi^{(I)}(t) = f_{(i,t)}\left[U^{0,t}_{i,t}U^{0,t+\Delta t}_{0,t}\left(U^{0,t+\Delta t}_{i,t+\Delta t}\right)^{-1}\left(U^{i,t+\Delta t}_{i,t}\right)^{-1} - 1\right] \tag{5.28}$$

If $f^{I}_{(i,t+\Delta t)} < f^{II}_{(i,t+\Delta t)}$, an arbitrage strategy would be to borrow $f_{(0,t)}$ amounts of cash and apply the second trading route. Then, at $t+\Delta t$, one would gain $f_{(0,t)}\left(U^{0,t}_{i,t}\right)^{-1}U^{i,t+\Delta t}_{i,t}U^{0,t+\Delta t}_{i,t+\Delta t}$ amounts of cash and owe $f_{(0,t)}U^{0,t+\Delta t}_{0,t}$, that is, the amount of cash with accrued interest. Then, one profits:

$$\Pi^{(II)}(t+\Delta t) = f_{(0,t)}\left[\left(U^{0,t}_{i,t}\right)^{-1}U^{i,t+\Delta t}_{i,t}U^{0,t+\Delta t}_{i,t+\Delta t} - U^{0,t+\Delta t}_{0,t}\right] \tag{5.29}$$

Since we're evaluating in terms of asset i, we have:

$$\Pi^{(II)}(t+\Delta t) = f_{(i,t)} \left[U_{i,t}^{i,t+\Delta t} U_{i,t+\Delta t}^{0,t+\Delta t} - U_{i,t}^{0,t} U_{0,t}^{0,t+\Delta t} \right] \tag{5.30}$$

which, discounted to time t, is equal to:

$$\Pi^{(II)}(t) = f_{(i,t)} \left[U_{i,t}^{i,t+\Delta t} U_{i,t+\Delta t}^{0,t+\Delta t} \left(U_{0,t}^{0,t+\Delta t} \right)^{-1} \left(U_{i,t}^{0,t} \right)^{-1} - 1 \right] \tag{5.31}$$

Because we don't know, *a priori*, which route will be used to exploit an arbitrage opportunity, the sum of both $\Pi^{(I)}(t)$ and $\Pi^{(II)}(t)$ is considered:

$$\Pi(t) = f_{(i,t)} \left[R(\gamma) + (R(\gamma))^{-1} - 2 \right] \tag{5.32}$$

where $R(\gamma) = U_{i,t}^{0,t} U_{0,t}^{0,t+\Delta t} \left(U_{i,t+\Delta t}^{0,t+\Delta t} \right)^{-1} \left(U_{i,t}^{i,t+\Delta t} \right)^{-1}$. If the quantity in rectangular brackets is zero, it means that there is no arbitrage opportunity. Here, we expose the risk-free profit quantity above as having a geometric interpretation: it is the curvature $R(\gamma)$ of the fibre bundle, introduced in Section 2.6 [21]. This quantity functions as a measure of curvature, which ties quite nicely with the idea of existent arbitrage, since the latter reflects market anomalies, associated with non-equilibrium.

Now we introduce a key observation, put forth by Ilinski (2000): there are superfluous degrees of freedom for each asset, regarding their units. That is, arbitrary changes in the units used to measure the asset's value should affect in no way the real, underlying dynamics: gauge symmetries are present in financial markets. Examples would be the insignificance of changing every one euro to one-hundred cents, splitting shares or trading in different currencies. None of this matters as long as there exists an understanding of the necessary scaling (Ilinski, 2000) [22].

In fact, consider an amount f_x of the asset x. Let us apply a local transformation to the fibre element and G-connection:

$$\begin{aligned} f_x &\xrightarrow{q} f'_x = f_x \cdot q(x) \\ U_x^y &\xrightarrow{q} U_x'^y = q(x) U(x,y) q^{-1}(y) \end{aligned} \tag{5.33}$$

[21] The change of notation relative to the one used in Chapter 2 is, once again, justified by the desire to keep consistent with the reference's notation.

[22] In a real-world scenario, this may have impacts, since, for example, if a company decides to split its shares, investors may extract some implicit information, affecting their behavior. This is what is called in physics a gauge symmetry breaking. It can be dealt with by adding perturbations to the ideal case, in order to approximate it to reality. Still, the latter is of utmost importance in market modelling (Ilinski, 2000).

where $q(x),\ q(y) \in G = \mathbb{R}_+$. The curvature, in the mathematical theory of differential geometry, should be invariant under these transformations. The local coordinates on which we operate should have no influence on the geometric, intrinsic characteristics of the underlying structure. The same occurs in the financial setting. Under our transformations – gauge transformation – let us analyse what happens to the curvature tensor, one of the constituents of the risk-free profit quantity previously discussed:

$$\begin{aligned} R(\gamma) \xrightarrow{q} R'(\gamma) &= q(x)R(\gamma)q^{-1}(x) = \\ &= R(\gamma) \end{aligned} \tag{5.34}$$

the same happens to the curvature tensor of opposite orientation. Hence, the gain or loss associated with the arbitrage opportunity is invariant under the gauge transformation.

With this in mind, the following is postulated: real world economic and financial phenomena are invariant under gauge transformations – local *numéraire* changes. Given that arbitrage gains around a closed trading path exhibit such invariance, they must be able to describe the market's dynamics, especially in non-equilibrium scenarios, given that arbitrage is associated with some market "malfunction", and, geometrically speaking, the implied existence of curvature paints a suggestive picture of deformation, perturbing the market. As we've stated before, the existent gauge-symmetry has to do with the indifference investors' show with regards to a reestablishment of measurement units.

5.2.2 Theoretical Extensions

Morisawa (2009) proposed a slight variation on the framework just described. He considers a space of N exchangeable assets – portfolio space, $\mathbb{R}^N$– and a trading strategy that involves exchanging in a closed loop. The existence of arbitrage opportunities is assumed (in the case of three currencies, it would be triangular arbitrage).

By trading in such a way to exploit this opportunity, beginning and ending the trade at the same portfolio, a gain or loss will occur. In other words, the trade won't be a closed path, even though we begin and end with the same kind of assets, and in the same proportions. To work in a closed path, he considered instead the portfolio ratio space. By picking any point in the portfolio space, one holds, in this case, some amount of the N assets. These amounts are exactly the coordinates of the chosen point in the portfolio space. However, if one multiplies each coordinate by the same scalar, one obtains exactly the same portfolio as before, only in a greater or smaller quantity. The nature of the portfolio is the same for every point in the line defined by this multiplication. We now introduce the notion of equivalence relation. An equivalence relation $\sim$ on $\mathbb{R}^N$ is a subset of $\mathbb{R}^N \times \mathbb{R}^N$, such that, for any element $a,\ b,\ c \in \mathbb{R}^N$, three conditions are respected:

1. $a \sim a$ (Reflexivity)

2. If $a \sim b$, then $b \sim a$ (Symmetry)

3. If $a \sim b$ and $b \sim c$, then $a \sim c$ (Transitivity)

We let the equivalence relation be: for any $a,\ b \in \mathbb{R}^N$, $a \sim b$ if $a = \lambda b,\ \forall \lambda \in \mathbb{R}_+$.

This defines the set $[a] = \left\{ b | b \sim a,\ \forall b \in \mathbb{R}^N \right\}$, called an equivalence class – the set of all elements equivalent with a. In turn, the set of all equivalence classes with this equivalence relation determines the space to be chosen – the real projective space:

$$\mathbb{R}\mathrm{P}^{N-1} = \left\{ [a] | \forall a \in \mathbb{R}^N \right\} \tag{5.35}$$

which has as elements every possible unique portfolio, regardless of quantities. Only the proportions matter. Hence the name portfolio ratio space. This way, starting with some portfolio $[x]$, and performing a trade around a closed path, effectively ending at the same portfolio, it doesn't matter if there was an arbitrage opportunity, since $[x] = \alpha [x]$, for all $\alpha \in \mathbb{R}$ – the trade is along a closed path when one works in the real projective space.

A fibre bundle structure is subsequently constructed by letting the base space be the real projective space $\mathbb{R}\mathrm{P}^{N-1}$ and the freedom associated with the scalar $\alpha \in \mathbb{R}$ define the fibre. What is obtained is the trivial fibre bundle $\mathbb{R}\mathrm{P}^{N-1} \times \mathbb{R}$.

As before, the structural group G acting on the fibre is the fibre itself, so we have a principal fibre bundle. The rest of the construction is exactly the same: the G-connections are prices and/or exchange rates, trading an amount of some portfolio for another is the parallel transport mediated by the connection, and the curvature of the connection is the variation of $\alpha \in \mathbb{R}$ under parallel transport along a closed path – the gain or loss of an arbitrage opportunity. The gauge symmetry is, again, the redenomination of asset's units.

The usage of the real projective space to explain market reality stops here, even though it has been suggested to provide the proper setting for financial modelling (Piotrowski & Sladkowski, 2006), given that most of economic and financial reality depends on proportions instead of nominal factors. That is to say, any indicator, whether it is prices or anything else, functions in relation to some other. There is a relativity present in financial economic reality.

Also, modifications were provided by Zhou & Xiao (2010), extending some of the ideas to option pricing. Their insight lies on the observation that option prices should have no dependency on *numéraire* changes – gauge transformations in the framework to be explained.

A financial market of N risky assets and a riskless one, such as a bond, serves as background. The vector $X = \left(x^1, x^2, ..., x^N\right)$ denotes the price of all assets, including the bond x^0, and $\mathrm{P} = \mathbb{R}_+^{N+1}$ is the set of all possible asset prices. Let the trivial manifold $M = \mathrm{P} \times \mathbb{R}$ be the price-time space, where $\mathbb{R}$ is the time axis.

The price dynamics are assumed to be driven by Brownian motions, which reflect the market's uncertainty, modelled by Itô processes:

$$dx^0 = x^0 \left(r(X,t)\, dt + \sum_{k=1}^{N} b_k^0 (X,t)\, dW^k \right) \tag{5.36}$$

$$\begin{cases} dx^i = x^i \left(\mu (X,t)\, dt + \sigma^i (X,t)\, dW^i\right) \\ dW^i dW^j = \rho^{ij} (X,t) \end{cases} \tag{5.37}$$

where $r(X,t)$ and $\mu(X,t)$ are drift terms, $\sigma(X,t)$ is the volatility and $\rho^{ij}(X,t)$ is the covariance matrix.

Then, a European option is considered, whose underlying assets are all of those belonging to the market. Let $V(X,t)$ be the value of the option at t. By value, it's meant the payoff that it would provide at that specific time.

To determine the dynamics, the Itô rule is followed:

$$dV = \left(\partial_t V + \frac{1}{2} \sum_{i,j=0}^{N} \rho^{ij} \sigma^i \sigma^j x^i x^j \partial_i \partial_j V \right) dt + \sum_{i=0}^{N} \partial_i V dx^i \tag{5.38}$$

It is standard to assume that a portfolio of the underlying assets with certain weights replicates the price dynamics of the option. This portfolio must be, however, self-financing, meaning that changes in portfolio's value depends solely on price changes. Let $F = \sum_{i=0}^{N} \partial_i V x^i$ be the mentioned portfolio, such that $dF = \sum_{i=0}^{N} \partial_i V dx^i$. Since it mimics the option's value, it's expected that $dV = dF$, and therefore:

$$\partial_t V + \frac{1}{2} \sum_{i,j=0}^{N} \rho^{ij} \sigma^i \sigma^j x^i x^j \partial_i \partial_j V = 0 \tag{5.39}$$

Let $\Omega = \frac{1}{2} \sum_{i,j=0}^{N} \rho^{ij} \sigma^i \sigma^j x^i x^j \partial_i \partial_j$. Then $(\partial_t + \Omega)\, V = 0$ completely describes the price evolution of the European option.

The problem with this equation is the lack of covariance under local *numéraire* changes of the underlying assets. If the price of some arbitrary asset is decided to be used as reference value for all other assets, it would be expected that the equation dictating the option's price dynamics remains invariant, given that the underlying asset prices are still the same and evolve in the same way as before. Notice that it's expected the price of the option to be different, but not the way that it evolves. Let $z(X,t)$ be the price of the asset to be used as *numéraire*. Under the transformation $V \to V' = z \cdot V$, the "new" pricing expression is different from $(\partial_t + \Omega) V = 0$ [23]. To deal with this, the authors made use of the fibre bundle structure and gauge theory.

The option price can be thought of as a function $V_i : M \to \mathbb{R}$, where $i \in \mathbb{N}$ is an index denoting all possible option prices for the same underlying assets, a set denominated $E_X = \{V_i | i \in \mathbb{N}\}$. This can also be seen as a price field, since it associates to each element of the space a single price. Employing the differential geometric formalism presented in Chapter 2, let the price-time space M be the base space and the set of all possible prices for the option, E_X, be the fibre. Then we have a fibre bundle geometry describing the model, where the fibre bundle is the disjoint union of all fibres $\mathrm{E} = \bigsqcup_{X \in M} E_X$. The set of possible *numéraire* changes is the structure group acting on the fibre, and a section of the fibre bundle is interpreted as the pricing of a specific option.

Returning to equation (5.39), let $g^{ij} = \frac{1}{2}\rho^{ij}\sigma^i\sigma^j x^i x^j \partial_i \partial_j$. These are elements of a positive definite matrix, which defines a metric for the space, turning M into a Riemannian manifold (essentially, a manifold endowed with a map – the metric – that allows for the notion of distance between points).

Firstly, the diffusion process presented in equation (5.39) has to be altered to a more general case, where this geometry is respected. This is achieved via the Laplace-Beltrami operator:

$$\Delta_g = |g|^{-\frac{1}{2}} \sum_{i,j=0}^{N} \partial_i \left(|g|^{\frac{1}{2}} g^{ij} \partial_j \right) \tag{5.40}$$

where $|g|$ is the determinant of the metric. This is a generalized version of the Laplacian operator for Riemannian manifolds. Hence, we obtain $(\partial_t + \Delta_g) V = 0$, the general version of equation (5.39).

Secondly, the derivative used has to be changed to a covariant one, where the fibre bundle structure of the market is also respected. This is achieved via the transformation:

$$\begin{aligned} \partial_i &\longrightarrow \nabla_i = \partial_i + A_i \\ \partial_t &\longrightarrow \nabla_t = \partial_t + B \end{aligned} \tag{5.41}$$

[23] Once again, order of application of group elements is meaningless here: all quantities commute. Hence, from now on, it is ignored, unless otherwise stated.

where A_i and B are the connection coefficients, $\forall i = 0, \ldots, N$. This leads to the covariant Laplace-Beltrami operator:

$$\tilde{\Delta}_g = |g|^{-\frac{1}{2}} \sum_{i,j=0}^{N} \nabla_i \left(|g|^{\frac{1}{2}} g^{ij} \nabla_j \right) \tag{5.42}$$

and the generalization of equation (5.39):

$$\left(\nabla_t + \tilde{\Delta}_g \right) V = 0 \tag{5.43}$$

For equation (5.43) to be covariant, it's necessary that under the gauge transformation $V \to V' = z \cdot V$, the option's value evolution remains invariant. Due to equation (5.41), the gauge transformation of the coefficients presented in (5.42) are known, and it can be proven that the evolution equation (5.43) is invariant:

$$\begin{aligned} &\left(\nabla_t + \tilde{\Delta}_g \right) V = 0 \longrightarrow z \cdot \left(\nabla_t + \tilde{\Delta}_g \right) V = 0 \Longrightarrow \\ &\Longrightarrow \left(\nabla_t + \tilde{\Delta}_g \right) V = \left(\nabla'_t + \tilde{\Delta}'_g \right) V \end{aligned} \tag{5.44}$$

Thus we obtain a covariant pricing equation. Finally, a condition to ensure that the fibre bundle is flat was calculated. Recalling expression (2.29), if $dA = 0$, then the fibre bundle is flat. This flatness is interpreted as the non-existence of arbitrage opportunities. The following expression is obtained:

$$\partial_l A_i = -\frac{1}{2} \partial_l \left\{ |g|^{-\frac{1}{2}} \sum_{j=0}^{N} \left[g_{ij} \sum_{k=0}^{N} \partial_k \left(|g|^{\frac{1}{2}} |g|^{kj} \right) \right] \right\} \tag{5.45}$$

where g_{ij} expresses the correlations in the market. If the above quantity is non-zero, there may be perturbations in the market allowing for arbitrage opportunities.

5.2.3 Modelling Dynamics

This idea of gauge symmetry existent in financial markets introduced in subsection 5.2 is a solely theoretical consideration. Regarding the formalism presented in 5.2, some numerical simulations were performed, initially by Ilinski (2001), in order to compare with real historical data, and with other pre-existent models. Further simulations were performed by authors such as Dupoyet et al. (2010, 2012) and Paolinelli & Arioli (2018, 2019), with the additional incorporation of some original ideas and variations, to be explained in subsection 5.2.3.2. We begin by providing a basic report of the base formalism used in the application of gauge theory to the modelling of market dynamics, expanded in Ilinski (2001).

5.2.3.1 Preliminaries

Let us return to Figure 5.4. Each node denotes a pair of asset and time, connected via links which represent the ability to trade. These trades are mediated via connections, acting on sections of the financial fibre bundle structure, which are simply nominal quantities that an investor may hold of that asset. We've observed that gains associated with arbitrage strategies around *plaquettes*, are invariant under local *numéraire* changes, that is, gauge invariant.

To build a model that describes the dynamics of prices and discount factors, that is, the gauge field in this framework, Ilinski (2001) proposed the following assumptions, here briefly described:

1. Properties of financial markets that govern dynamics must be gauge-invariant;
2. Financial markets are intrinsically uncertain;
3. Existence of arbitrage opportunities is allowed, but minimized;
4. For a complete financial market, dynamics are local;
5. Results from portfolio theory and derivative pricing must be reproduced in the "classical" limit.

In order to achieve this, a gauge-invariant functional, denoted as s_{Gauge} [24], known as the action, is stipulated to be responsible for any gauge-invariant dynamics of the system. It is constructed, for a general case, by considering the risk-free profit of arbitrage strategies associated with all possible elementary closed trading path – *plaquettes*. As we've seen, this amounts to the product of connection operators for *plaquettes*, which can be thought of in terms of curvature tensor elements. Hence, we have:

$$s_1\left(\{U_\gamma\}\right) = \sum_{\{\gamma_1,\gamma_2,\dots,\gamma_n\}} \alpha_{\gamma_1,\gamma_2,\dots,\gamma_n} R(\gamma_1)R(\gamma_2)\dots R(\gamma_n) \tag{5.46}$$

where $\{\gamma_1,\gamma_2,\dots,\gamma_n\}$ is a collection of all possible sets of elementary *plaquettes*, $\alpha_{\gamma_1,\gamma_2,\dots,\gamma_n} \in \mathbb{R}$ are arbitrary coefficients meant to represent volatility and $R(\gamma) = \prod_{\gamma\in\{\gamma_1,\gamma_2,\dots,\gamma_n\}} U_\gamma - 1$. [25] Here, it is generalized to any *plaquette* of any size.

[24] From now on, we'll denoted this action as s_1. By doing this, we foreshadow the construction of other actions.

[25] This product is interpreted as the product of all connection operators associated with each portion of the loop γ.

This is only possible, of course, with the assumption that these "rectangular" arbitrage trading paths function as the building blocks for more complicated strategies. The probability of observing a certain system configuration, that is, a certain price and rate of return, is proportional to the exponential of that action:

$$P\left(\{U_\gamma\}\right) \propto \mathrm{e}^{-s_1\left(\{U_\gamma\}\right)} \Longrightarrow \ln(P\left(\{U_\gamma\}\right)) = -s_1\left(\{U_\gamma\}\right) \tag{5.47}$$

where the equality is verified up to an additive constant. Let us consider the simple case of only two assets – cash and a generic share – and a linear action [26], that is:

$$s_1\left(\{U_\gamma\}\right) = \sum_{\gamma\in\{\gamma_1,\gamma_2,\ldots,\gamma_n\}} \alpha_\gamma R(\gamma) \tag{5.48}$$

which can be further simplified by noticing that, per *plaquette*, we have two possible paths, different only in orientation. Therefore, we obtain:

$$\begin{aligned} s_1\left(\{U_p\}\right) &= \sum_{p\in\{plaquettes\}} \alpha_p\left[R(p)+R^{-1}(p)\right] = \\ &= \sum_{p\in\{plaquettes\}} \alpha_p\left[\prod_p U_p + \prod_p U_p^{-1} - 2\right] \end{aligned} \tag{5.49}$$

where $p \in \{plaquettes\}$ is a possible *plaquette*. Since we are considering a two asset system, it is clear that the amount of *plaquettes* is related to the time horizon in this ladder geometry. For this reason, we'll substitute the index p with the time variable t, after being discretized as $t_j = j\Delta t$, for $j = 0,\ 1,\ \ldots,\ l$, where Δt is the smallest time frame possible and $T = l\Delta t$ is the time horizon.

Instead of working with the parallel transport operators U_p, let S_j be the price of a share in units of cash at time t_j, and let $e^{r_0(t)\Delta t}$ and $e^{r_1(t)\Delta t}$ be the rates of return of cash (the 0^{th} asset) and share (the 1^{st} asset), respectively. These are the connections which constitute the curvature tensor, which in turn make up the action. Recalling that the action is gauge-invariant, we can simply fix connection values, which change in no way the dynamics, but simplify the calculations (this is known as gauge fixing). The connections linking nodes in the time direction will be fixed to $e^{r_0\Delta t}$ and $e^{r_1\Delta t}$, and the one in the asset direction will be fixed at time $t_0 = t$ to S_0. Therefore, by replacing the parallel transport operators in equation (5.49) with the new values, we obtain:

$$s_1\left(\{S_j\}\right) = \sum_{j=0}^{l-1} \alpha_j\left(S_j^{-1}e^{r_1\Delta t}S_{j+1}e^{-r_0\Delta t} + S_j e^{r_0\Delta t}S_{j+1}^{-1}e^{-r_1\Delta t} - 2\right) \tag{5.50}$$

[26] Correlation between assets could be introduced by considering actions other than linear (Ilinski, 2001).

which in the continuous limit $\Delta \to 0$ converges to:

$$s_1(S_t) = \frac{1}{2\sigma^2}\int_0^T (\partial_t \ln S_t - r)^2 \, dt \tag{5.51}$$

with $\alpha_t \equiv \alpha = 1/\left(2\sigma^2\right)$, where σ^2 is the volatility[27], and $r \equiv r_0 - r_1$, which is the average rate of share return. Returning to equation (5.47) we see that the probability of a configuration being characterized by the price S_t (in a continuous time case) and average rate of share return r is:

$$P(S_t) \propto \exp\left\{-\frac{1}{2\sigma^2}\int_0^T (\partial_t \ln S_t - r)^2 \, dt\right\} \tag{5.52}$$

To understand how prices vary from, let's say, time 0 to time T, we must consider the transition probabilities associated to the price at each time. This can be done with the notion of path-integrals. Very succinctly, infinitesimal changes in prices can be seen as infinitesimal trajectories in price space, dS_t. The total change of this quantity would be evaluated by an integral such as:

$$\int_0^{+\infty} \frac{dS_t}{S_t} = \int_0^{+\infty} d\ln S_t \tag{5.53}$$

If price were to be interpreted as a stochastic process, this would return the classical value, which is the Itô integral of a stochastic process[28]. Instead of this approach, let us associate to price trajectory at each time, a probability weight obtained via the action previously considered, calculated for each *plaquette*, which, as stated before, is associated to time. Therefore, we get the following transition probability:

$$P(S_T|S_0) \propto \int e^{-s_1} D\ln S \tag{5.54}$$

where and $D\ln S$ denotes integration over all possible price paths, that is:

$$\int D\ln S = \prod_{j=1}^{l-1} \int_0^{+\infty} d\ln S_j \tag{5.55}$$

where $\ln S_j$ denotes the log-price at the j^{th} time step[29].

[27] The time dependence of the volatility term was dropped for simplicity, at the cost of generality and realism. However, for high-frequency trading, this simplification doesn't present any major drawbacks (Dupoyet et al. 2012).

[28] In financial literature, the price dynamics are described by stochastic process $dS_t = S_t\left(r_{1,t}dt + \sigma_t dW_t\right)$, where W_t are Brownian motions, respectively. Furthermore, the measure of integration is dS_t/S_t instead of dS_t, because the former is gauge-invariant, while the latter is not, as long as the gauge group associated is the group of dilations (the derivative of a constant is zero).

[29] Sometimes the logarithm is used instead of the natural logarithm, but since the difference between both is just a multiplicative constant, these differences aren't of fundamental importance.

This notion of an "action" is completely borrowed from physics, and it actually provides a surprisingly intuitive idea for its need in finance. In physics, this quantity is minimized in order to derive the equations of motion. Classically, particles follow paths of least resistance in their movement, represented by the minimization of the action. However, when one considers instead a quantum system, probabilities associated to each dynamical variable must be considered. To introduce randomness, one considers, instead of a single path, an infinite number of them, where the most probable one is the "classical" path. This is captured by an equation similar in form to (5.55). This idea was introduced in physics by Feynman (1948), and bears the name of path-integrals. Here, it is suggested a similar approach. Essentially, for each trajectory in price space, a probability is associated. For example, if for a particular price trajectory the possible gains of an arbitrage strategy associated to that *plaquette* are very high (low), the exponential will be very low (high), and therefore, that price trajectory will be very unlikely. As such, the most probable price path is that where arbitrage is zero, thus satisfying one of the initial axioms – arbitrage is allowed but minimized. Integral (5.54) can be evaluated, which gives the familiar result of prices following a log-normal distribution. We've obtained the "classical" result, satisfying the 5^{th} assumption.

These considerations provide the dynamics of the gauge field, that is, of prices at each point in the lattice. Without economic agents interacting with the field, a financial market would be composed of assets whose prices followed such a distribution. However, the way investor holdings interact with each other via the gauge field – the effect of the gauge field on the trading agents – must be taken into account. In order to do this, the following assumption (among others) is key: An investor starting at some point x_0 in the lattice of only two assets, wants to maximize wealth, with the available information, by following some trading path γ, until x_n:

$$s_2(\gamma) = \ln\left(f_{x_n}^{-1}\left(\prod_{i=1}^{n} U_{x_i}\right) f_{x_0}\right) \tag{5.56}$$

where U_{x_i} may be prices or interest rates associated with the path which starts at x_{i-1} and ends at x_i. It depends on what the investor perceives to be the best strategy[30]. Notice that the indices in x denote some time step, at which the investor holds some amount of the asset, whether it is cash or share. In other words, x_0 and x_n represent some asset (cash or share) at the beginning and end of the trading strategy, respectively.

[30] f_{x_0} and $f_{x_n}^{-1}$ are elements of the fibre. By applying them to the product of several connection terms, we are essentially turning it into a gauge-invariant quantity (notice that before, the product of many connections was said to be gauge-invariant, but in that case we were considering loops, not arbitrary paths), allowing for the comparison with other paths. Financially speaking, the investor starts with an amount f_{x_0} of asset x_0 and by following this trading path, ends up with a certain amount of asset x_n, that can be expressed unit-free by applying $f_{x_n}^{-1}$ (this quantity is not necessarily equal to the inverse of the gains accrued by following the investment path).

We can interpret this action as describing the gains an investor accrues by following a trading path which is perceived to be optimal. The logarithm is necessary to make sure that maximizing in each step of the transaction maximizes the whole trajectory.

Therefore, we may assume that an investor follows some path γ with probability:

$$P(\gamma) \propto \exp\left\{\tilde{\beta} \cdot s(\gamma)\right\} = f_{x_n}^{-\tilde{\beta}} \left(\prod_{i=1}^{n} U_{x_i}^{\tilde{\beta}}\right) f_{x_0}^{\tilde{\beta}} \tag{5.57}$$

where $\tilde{\beta}^{-1}$ represents the average bounded rationality of investors per unit time (Ilinski, 2001). To model the investor's behaviour, we assume that at some time t_j, there are probabilities of the investor holding cash or shares, $\mathrm{p}_{0,j}$ and $\mathrm{p}_{1,j}$, respectively. For an infinitesimal time step Δt, the expression above functions as a transition probability from the current state to the future one. The quantity in parenthesis functions as an evolution operator. In matrix form, the transition probability is[31]:

$$\mathrm{P}\left(t_{j+1} : t_j\right) = \begin{pmatrix} f_{0,j+1}^{-\tilde{\beta}} & 0 \\ 0 & f_{1,j+1}^{-\tilde{\beta}} \end{pmatrix} \begin{pmatrix} e^{\tilde{\beta} r_0 \Delta t} & S_{j+1}^{\tilde{\beta}} \\ S_{j+1}^{-\tilde{\beta}} & e^{\tilde{\beta} r_1 \Delta t} \end{pmatrix} \begin{pmatrix} f_{0,j}^{\tilde{\beta}} & 0 \\ 0 & f_{1,j}^{\tilde{\beta}} \end{pmatrix} \tag{5.58}$$

By fixing the gauge as $f_{0,j} = S_j^{1/2}$ and $f_{1,j} = S_j^{-1/2}$ we have:

$$\begin{cases} \mathrm{p}_{0,j+1} = e^{\tilde{\beta} r_0 \Delta t - \frac{\tilde{\beta} \ln(S_{j+1}/S_j)}{2}} \mathrm{p}_{0,j} + e^{\frac{\tilde{\beta} \ln(S_{j+1}/S_j)}{2}} \mathrm{p}_{1,j} \\ \\ \mathrm{p}_{1,j+1} = e^{-\frac{\tilde{\beta} \ln(S_{j+1}/S_j)}{2}} \mathrm{p}_{0,j} + e^{\tilde{\beta} r_1 \Delta t + \frac{\tilde{\beta} \ln(S_{j+1}/S_j)}{2}} \mathrm{p}_{1,j} \end{cases} \tag{5.59}$$

An interpretation is provided: if interests associated to a cash account are high and the relative price of a share is also high, the probability of the investor being in a cash account increases, since the gains are favourable, and buying shares is not worthwhile; the same rationale can be applied to the probability of holding a share.

Now, in order to generalize to many investors, considerations will shift from each investor itself to the amount of cash and share units available in the market at some time t_j. Let them be denoted by n_j and m_j, respectively. Changes to these amounts will be modelled with the use of operators $\hat{\psi}_{0,j}^{+}$ and $\hat{\psi}_{1,j}^{+}$, responsible for the creation, and operators $\hat{\psi}_{0,j}^{-}$ and $\hat{\psi}_{1,j}^{-}$, responsible for the annihilation, of a unit of cash or share, respectively. At each time t_j, we can attribute to the market a two dimensional vector denoting the amount of cash

[31] The author also takes into account transaction costs, which we ignore for ease of exposition. Later applications ignore it as well. However, transaction costs play an important role in the dynamics of the system, as a stabilizer, effectively stopping small mispricing to have a significant effect on money flows (Ilinski, 2001).

and shares, which in turn manifest how many investors are in each asset. These operators act exactly on those state vectors, and produce a quantity which effectively labels the state of the market in terms of the portfolio (this quantity is known as eigenvalue, and the vector is known as eigenvector). This way, the behaviour of many investors is considered not by their singular actions, but by the effect they have on the amounts of cash and share inside the market[32].

After some mathematical considerations, and also the employment of some machinery from quantum mechanics and quantum field theory (for a derivation and discussions see Ilinski (1997), Ilinski (2001), Sokolov et al. (2010) and Paolinelli & Arioli (2018)), the following action is obtained:

$$\begin{aligned} s_2 = \frac{1}{\tilde{\beta}} \sum_{j=0}^{l} \Bigg[& \Big(\left(\bar{\psi}_{0,j+1} - \bar{\psi}_{0,j} \right) \psi_{0,j} \Big) \\ & + \Big(\left(\bar{\psi}_{1,j+1} - \bar{\psi}_{1,j} \right) \psi_{1,j} \Big) \\ & + \Big(\bar{\psi}_{0,j+1} e^{-\tilde{\beta} r t_j} e^{-\tilde{\beta} r_0 \Delta t} S_j^{\tilde{\beta}} \psi_{1,j} \Big) \\ & + \Big(\bar{\psi}_{1,j+1} e^{\tilde{\beta} r t_j} e^{-\tilde{\beta} r_1 \Delta t} S_j^{-\tilde{\beta}} \psi_{0,j} \Big) \Bigg] \end{aligned} \tag{5.60}$$

We present interpretations for each bracket, in order:

1. $\left(\bar{\psi}_{0,j+1} - \bar{\psi}_{0,j} \right) \psi_{0,j}$ denotes relative changes in the amount of cash during a time step;

2. $\left(\bar{\psi}_{1,j+1} - \bar{\psi}_{1,j} \right) \psi_{1,j}$ denotes relative changes in the amount of shares during a time step;

3. $\bar{\psi}_{0,j+1} e^{-\tilde{\beta} r t_j} e^{-\tilde{\beta} r_0 \Delta t} S_j^{\tilde{\beta}} \psi_{1,j}$ compares the amount of cash at time t_{j+1} with the amount of cash at time t_j resultant from investors selling shares at price $S_j^{\tilde{\beta}}$;

4. $\bar{\psi}_{1,j+1} e^{\tilde{\beta} r t_j} e^{-\tilde{\beta} r_1 \Delta t} S_j^{-\tilde{\beta}} \psi_{0,j}$ compares the amount of shares at time t_{j+1} with the amount of shares at time t_j resultant from investors selling cash at price $S_j^{-\tilde{\beta}}$ (that is, buying shares at price $S_j^{\tilde{\beta}}$).

Therefore, we may think of the action (5.60) as the "flow" between shares and cash characterizing a particular lattice-like financial market, associated with the gains and

[32] Two important details: First, we still refer to investors, even though cash and share amounts are being considered, and second, a closed market is assumed, that is, $n_j + m_j = M$, for all t_j, where M denotes the total number of traded lots, both cash and shares (Ilinski, 2001). Because the notion of investors is tied together with their trading behavior, M can also be interpreted as the total number of investors (Paolinelli & Arioli, 2018).

losses of the investors operating in the market. Then, an expression for an evolution operator connecting two portfolio states at two different times is obtained:

$$I[\psi,\bar{\psi}] = \int e^{\tilde{\beta} s_2} D\psi_0 D\bar{\psi}_0 D\psi_1 D\bar{\psi}_1 \tag{5.61}$$

which can then be used to construct a transition probability of the portfolio allocation (n_j, m_j) at time t_j changing to (n_{j+1}, m_{j+1}) at time t_{j+1}, effectively generalizing the transition probability (5.58) to a many-investors scenario. The integration in $I[\psi,\bar{\psi}]$ is performed over all possible trajectories in portfolio space, some more probable than others, "decided" by the exponential of the action.

To introduce the first action, built with arbitrage gains around elementary *plaquettes*, we need only to use instead the following functional integral in the transition probability:

$$I[\ln S, \psi, \bar{\psi}] = \int e^{-s_1 + \tilde{\beta} s_2} D\ln S D\psi_0 D\bar{\psi}_0 D\psi_1 D\bar{\psi}_1 \tag{5.62}$$

Then, the transition probability denotes the probability of some transition in prices and portfolio allocations to occur.

Ilinski (2001) also provides a perturbation [33] to the action s_1. It represents how the investors themselves influence the gauge field. It is achieved by adding to the first action, inside the quadratic term, what is denoted as the Farmer's term:

$$F = \frac{\alpha}{M}\left(\bar{\psi}_0\psi_0 - \bar{\psi}_1\psi_1\right) \tag{5.63}$$

where α is some constant representing share liquidity (Paolinelli & Arioli, 2018). The Farmer's term allows for prices to rise (drop) when somebody buys (sells) the share.

The final action can be presented in discrete form as:

$$\begin{aligned} s &= -s_{1F} + s_2 = \\ &= -\frac{1}{2\sigma^2}\sum_{j=0}^{l}\left(\ln S_{j+1} - \ln S_j - r + \frac{\alpha}{M}\left(\bar{\psi}_0\psi_0 - \bar{\psi}_1\psi_1\right)\right)^2 + \\ &+ \frac{1}{\tilde{\beta}}\sum_{j=0}^{l}\left[\left(\left(\bar{\psi}_{0,j+1} - \bar{\psi}_{0,j}\right)\psi_{0,j}\right) + \left(\left(\bar{\psi}_{1,j+1} - \bar{\psi}_{1,j}\right)\psi_{1,j}\right) + \right. \\ &\left. + \left(\bar{\psi}_{0,j+1} e^{-\tilde{\beta} r t_j} e^{-\tilde{\beta} r_0 \Delta t} S_j^{\tilde{\beta}} \psi_{1,j}\right) + \left(\bar{\psi}_{1,j+1} e^{\tilde{\beta} r t_j} e^{-\tilde{\beta} r_1 \Delta t} S_j^{-\tilde{\beta}} \psi_{0,j}\right)\right] \end{aligned} \tag{5.64}$$

[33] Also, a second perturbation is introduced meant to mimic the tendency for investors to exhibit "herd" behavior (Ilinski, 2001).

The dynamics here presented inspired variations and applications (Dupoyet et al., 2010; Dupoyet et al., 2012; Paolinelli & Arioli, 2018; Paolinelli & Arioli, 2019). To be noticed that better agreement with historical data is verified with the subsequent applications, compared with this model. The latter fitted quite nicely with empirical data in the middle region, but failed to accurately reproduce the characteristic fat tails (Paolinelli & Arioli, 2018).

5.2.3.2 Numerical Simulations

Following the model proposed by Ilinski (2001), Dupoyet et al. (2010) explored the gauge-invariant lattice model for a financial market. The geometry they explore shares similarities with the previous one, in the sense that a generic asset and cash are connected, throughout time, in a ladder-like fashion, with the addition of more assets, all of them connected solely with cash.

Here, assets can be traded with one another by firstly exchanging with cash. A change in notation was also introduced by Dupoyet et al. (2010, 2012):

- To each node in the lattice, the pair (i, j) is attributed, where i denotes the asset and j the time;
- Each node in the lattice is populated by a field $\Phi(i, j)$, for all (i, j), representing the amounts of asset i held at time j (elements of the fibre associated to each pair (i, j));
- Each link between nodes is populated by a field $\Theta_k(i, j)$, for all (i, j) and $k = 0,\ 1$. If $k = 0$, then $\Theta_0(i, j)$ is interpreted as the conversion factor between (i, j) and $(i, j+1)$, that is, the interest rate of asset i. If $k = 1$, then $\Theta_1(i, j)$ is interpreted as the conversion factor between (i, j) and $(i+1, j)$, that is, the exchange rate between assets i and $i+1$. These fields are nothing but gauge fields, constituted by the parallel transport operators defined by the choice of connection.

Opposing the formalism by Ilinski (2000), here the parallel transport operators act on the fibre elements in the following way:

$$\begin{cases} \Phi(i, j) = \Theta_1(i, j)\Phi(i+1, j) \\ \Phi(i, j) = \Theta_0(i, j)\Phi(i, j+1) \end{cases} \tag{5.65}$$

that is, inverted versions of the connections previously considered in 5.2.1.

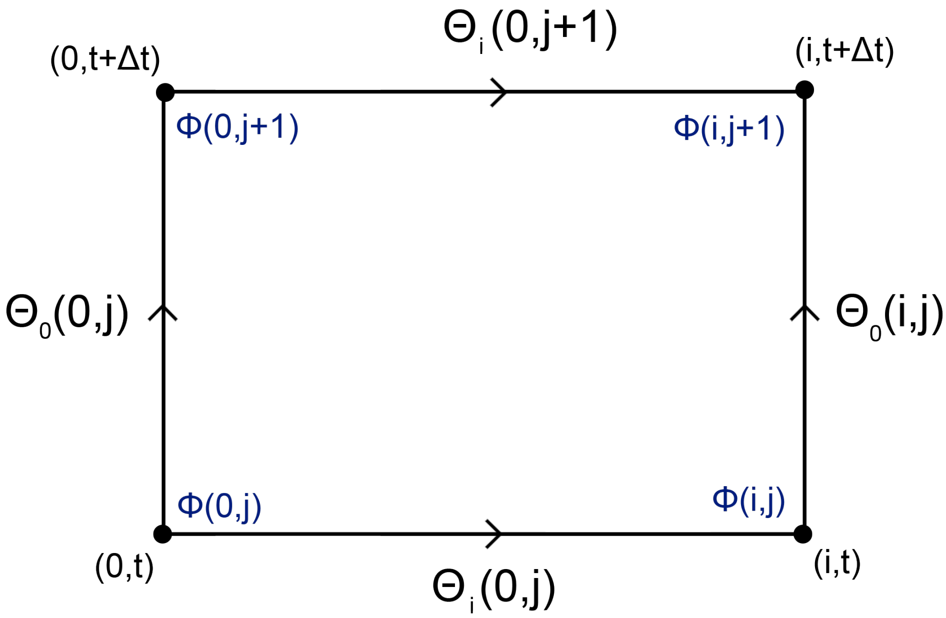
Because each asset can only be traded with cash, via the connection "price", we can consider each asset's elementary *plaquette* that it creates, at some time, with cash, individually (see Figure 5.6).

Once again, an action dependent of the gauge field is constructed with the arbitrage gains possible in the *plaquette*. Hence, we have:

$$s_1[\Theta] = \frac{1}{2}\sum_{j=0}^{l-1}\sum_{i=1}^{N}\left[R_{i,0} + R_{0,i} - 2\right] \tag{5.66}$$

where $R_{i,0} = \Theta_i(0,j)\Theta_0(i,j)\Theta_i^{-1}(0,j)\Theta_0^{-1}(0,j)$ and $R_{0,i} = R_{i,0}^{-1}$. There are two summations: one over all times, thus extending the arbitrage gains to all *plaquettes*, and one over all assets. If we substitute each gauge field element for its equivalent in the base formalism, that is, for prices and interest rates, we quickly observe that it is in fact the inverse of the previous formalism. Therefore, there is no difference in the gauge field dynamics between each.

Figure 5.6: Similar elementary *plaquette* to that shown in 5.2.1, with different notation and opposite direction.



Source: Dupoyet et al. (2010).

A second action is also introduced, which makes use of the covariant derivatives of the fibre elements. For each point (i,j), we can define two covariant derivatives acting on $\Phi(i,j)$: one calculating the "forward derivative", $\nabla^+\Phi$, and another calculating the "backward derivative", $\nabla^-\Phi$. The same can be done to $\bar{\Phi}(i,j) = \Phi^{-1}(i,j)$, thus returning the covariant derivatives $\bar{\nabla}^+\bar{\Phi}$ and $\bar{\nabla}^-\bar{\Phi}$. Finally, we must take in account the direction on which these covariant derivatives are being performed. Since in this model we are only considering rectangular *plaquettes*, there are only three possible directions (four sides of the rectangle minus one which is only "available" in the next time period). Therefore, this action is composed, for each point (i, j), of twelve quantities, which are covariant derivatives for all possible directions.

A financial interpretation of the covariant derivative is none other than the gains obtained in some transaction, measured in a common unit. To transform this into a gauge-invariant quantity, in order to be effectively used as a second action, instead of the covariant derivatives, the quantities $\bar{\Phi}\left(\nabla^{\pm}\Phi\right)$ and $\left(\bar{\nabla}^{\pm}\bar{\Phi}\right)\Phi$ are used. They can be interpreted as relative gains (with associated risk) obtained via transactions. To provide an expression for the action, let k be one of the three possible directions – cash between two time periods, asset between two time periods and cash to asset (or vice-versa) – which are denoted as *axis*, *time* and *space* directions, respectively. Therefore, $k \in \{axis,\ time,\ space\}$. Then, the second action is as follows[34]:

$$s_2\left[\Theta,\ \Phi,\ \bar{\Phi}\right] = \sum_x \sum_k \Big[d_k^+ \bar{\Phi}(x)\left(\nabla_k^+ \Phi(x)\right) + d_k^- \bar{\Phi}(x)\left(\nabla_k^- \Phi(x)\right) + \bar{d}_k^+ \left(\bar{\nabla}_k^+ \bar{\Phi}(x)\right)\Phi(x) + \bar{d}_k^- \left(\bar{\nabla}_k^- \bar{\Phi}(x)\right)\Phi(x)\Big] \tag{5.67}$$

where $d_k^{\pm},\ \bar{d}_k^{\pm} \in \mathbb{R}$ are constants which determine the interactions between the gauge field and the fields Φ and $\bar{\Phi}$. Also, with this notation, we can explicitly show the covariant derivatives:

$$\nabla_k^- \Phi(x) = \Theta(x)\,\Phi(x + e_k) - \Phi(x) \tag{5.68}$$

$$\nabla_k^- \Phi(x) = \Theta^{-1}(x - e_k)\,\Phi(x - e_k) - \Phi(x) \tag{5.69}$$

$$\bar{\nabla}_k^+ \bar{\Phi}(x) = \bar{\Phi}(x + e_k)\,\Theta^{-1}(x) - \bar{\Phi}(x) \tag{5.70}$$

$$\bar{\nabla}_k^- \bar{\Phi}(x) = \bar{\Phi}(x - e_k)\,\Theta(x - e_k) - \bar{\Phi}(x) \tag{5.71}$$

whereby e_k we mean to say that the unitary vector must be added to, in this case, point x, pertaining to the direction k. Finally, a partition function is built with the action $s = s_1 + s_2$. Then:

$$Z(\beta) = \int e^{-\beta \cdot s\left[\Theta,\ \Phi,\ \bar{\Phi}\right]} D\Phi D\Theta \tag{5.72}$$

[34] Even though $x \equiv (i, j)$ is an unnecessarily generalized way of denoting coordinates in the lattice, it permits us to write the action in a more compact form.

where β is a parameter with a similar interpretation as before. Once again, for all possible trajectories of the gauge field Θ and traded lots Φ, a path integral is calculated, with a probabilistic weight associated to it. Its financial interpretation is the following: arbitrage gains are allowed but minimized, as before, but also the possible gains and losses. With this, expected values of gauge-invariant quantities (observables of the financial market) can be estimated:

$$\langle O\rangle = Z^{-1}(\beta)\int O\left[\Theta,\ \Phi,\ \bar{\Phi}\right] e^{-\beta\cdot s\left[\Theta,\ \Phi,\ \bar{\Phi}\right]} D\Phi D\Theta \tag{5.73}$$

where $O\left[\Theta,\ \Phi,\ \bar{\Phi}\right]$ is some financial observable. It's nothing more than the weighted average. In practice, to realize the numerical simulation, the Dupoyet et al. (2010) decomposed the action into two components, a dependent and an independent of the field, for both fields, and expressed their value in exponential form. The following are probability density functions for $\theta_\mu(x)$ and $\phi(x)$:

$$p_\Theta\left(\theta_\mu(x)\right) \propto \exp\left[-\beta\left(\bar{L}_\Theta\, e^{\theta_\mu(x)} + e^{-\theta_\mu(x)} L_\Theta\right)\right] \tag{5.74}$$

$$p_\Phi\left(\phi(x)\right) \propto \exp\left[-\beta\left(\bar{L}_\Phi\, e^{\phi(x)} + e^{-\phi(x)} L_\Phi\right)\right] \tag{5.75}$$

where $\Theta_\mu(x) = \exp\left(\theta_\mu(x)\right)$ and $\Phi(x) = \exp\left(\phi(x)\right)$. $L_\Theta,\ \bar{L}_\Theta > 0$ and $L_\Phi,\ \bar{L}_\Phi > 0$ are dependent on the local environment, and independent of the fields at that point x.

To generate the ensemble of fields $[\Theta,\ \Phi]$, they begin by choosing them at random, and then, through an heat bath algorithm, the fields are locally updated with recourse to the probability densities exposed above. With this ensemble, expected values of financial variables can be calculated. The chosen observable was:

$$\Lambda_\ell(i,j) = \log W_\ell(i,j) = \log\left[\bar{\Phi}(i,j-\ell)\left(\prod_{k=1}^{\ell}\Theta_0(i,j-k)\right)\Phi(i,j)\right] \tag{5.76}$$

which denotes the logarithm of the relative profit made with asset i at time j, after holding it for a time interval of size ℓ. Notice the similarities with the second action (5.56) proposed by Ilinski (2001), presented in 5.2.3.1.

Calculating the expected value $\langle\Lambda_\ell(i,j)\rangle$ in this model takes into account any trading activity, even possible arbitrage opportunities. There are $N(l-\ell+1)$ measurements available of this expected value, forming the set $\Lambda = \{\langle\Lambda_\ell(i,j)\rangle\}$ [35]. In order to compare

[35] To prevent any confusion, l denotes the market time horizon, while ℓ denotes the investment time horizon, that is, the amount of time the investor holds on to the asset.

with historical data, which is available in the form of historical prices, some changes were introduced. First, instead of prices, $\log(S(t')/S(t))$ is used, where $t' > t$. This way the comparison with $\langle \Lambda_\ell(i,j) \rangle$ is possible. Second, historical prices are cumulative measures, and as such, reflect the various time lengths that each investor holds the asset, determined by the laws of supply and demand. In the lattice construction this effect is not present, and to remedy this, Dupoyet et al. (2010) proposed for each element of Λ to be selected with probability $\ell^{-\nu}$, where ν is a parameter, which, if positive, reflects the situation where fewer investors hold assets for longer periods of time. This way, relative gains associated to holding some asset for long periods of time are less probable to occur in the market. This probabilistic weight follows what is known as a power law, which is devoided of scale, and thus, compatible with a gauge-invariant formalism. This new stochastic variable will be referred to as X_L.

Regarding the numerical simulation, it was performed on a lattice of size $l = 260$, which is interpreted as time steps, and $N = 30$ assets. All parameters were fixed in an *ad hoc* fashion. Then the simulation will provide a distribution of the number of relative gains chosen from Λ, per small variation of X_L. Results were compared with historical data extracted from the NASDAQ index, with a sampling time interval of one minute, from a sample of close to two years. An agreement until four orders of magnitude of the lattice simulation (complemented with the holding time adjustment) with historical data is observed, although it underestimates slightly the probability of large market corrections in the fat tail region (Paolinelli & Arioli, 2018). However, it still implies that the gauge principle is verified to a certain extent.

Building upon the latter model, Dupoyet et al. (2012) introduced a different strategy for updating the lattice constituents. To explore its effects, a single asset (and cash) was considered in the simulations. Firstly, they did a gauge fixing for all fields pertaining to the "cash" axis, obtaining $\Phi(0,j) = \bar{\Phi}(0,j) = 1$, for all j, and $\bar{\Phi}(0,j)\Theta_0(0,j)\Phi(0,j+1) = W_1(0,j)$, that is, the return associated to the cash account is fixed (similar to what was done by Ilinski (2001)). However, the authors manifest a clear interest in modelling high-frequency trading, where they argue that the account interest rate is insignificant. For this reason, let $W_1(0,j) = 1$. These gauge fixings are performed at every step of the simulation. Secondly, they ran the previous algorithm in order to achieve an equilibrium situation [36]. This way, a financial market which allows the existence of arbitrage opportunities for brief moments, due to the uncertain nature of markets, is achieved. It is with this equilibrium configuration that the authors introduce the new

[36] By this, it is meant that the ensemble of fields started with random values, and through a Monte Carlo method, were replaced successively until it stabilized near an ensemble closely resembling what was to be expected from the probability densities shown before.

updating strategy. It permits the local perturbation of the fields, post-equilibrium. This local perturbation begins by considering the following quantity:

$$v_j = r_j(r_{j+1} - r_{j-1}) \tag{5.77}$$

where $r_j = \log W_1(1,j) = \log\left(\bar{\Phi}(1,j)\Theta_0(1,j)\Phi(1,j+1)\right)$, which is gauge-invariant by construction. Even though this quantity was introduced on empirical terms, it has also a theoretical justification. It is the discretized version of the derivative of the Wiener process which drives the dynamics of the returns (Dupoyet et al., 2011).

In the equilibrium configuration, v_j is calculated for each j, and when the maximum value, in absolute terms, is found, the field components associated to that site are updated.

At each site j, the fields $\Theta_0(1,j-2)$, $\Theta_0(1,j-1)$, $\Theta_0(1,j)$, $\Phi(1,j-1)$ and $\Phi(1,j)$ are the only ones updated. The update works by first selecting random values from the probability density functions:

$$p_\Theta\left(\theta_\mu(x)\right) \propto \exp\left[-2\beta\sqrt{L_\Theta \bar{L}_\Theta}\,\cosh\left(\theta_\mu(x)\right)\right] \tag{5.78}$$

$$p_\Phi\left(\phi(x)\right) \propto \exp\left[-2\beta\sqrt{L_\Phi \bar{L}_\Phi}\,\cosh\left(\phi(x)\right)\right] \tag{5.79}$$

which are the same probability density functions in (5.74) and (5.75), after specific gauge fixings (see Dupoyet et al. (2012)). Then, the following averages are calculated:

$$\begin{aligned} a_\theta &= \frac{1}{3}\sum_{j=j-2}^{j} \theta_0(1,k) \\ a_\phi &= \frac{1}{2}\sum_{j=j-1}^{j} \phi(1,k) \end{aligned} \tag{5.80}$$

and each field is updated accordingly:

$$\begin{aligned} \theta_0(1,k) &\longrightarrow \theta_0(1,k) - \chi a_\theta \\ \phi(1,k) &\longrightarrow \phi(1,k) - a_\phi \end{aligned} \tag{5.81}$$

where χ is a parameter. Finally, because these updates perturb the minimal-arbitrage environment initially achieved, then the heat bath algorithm is subsequently applied to the fields connecting asset with share, for each time j, that is, $\Theta_1(0,j)$, for $j \in [j-1,\ j+1]$. The simulation was performed on a lattice of size $l = 782$, and a single asset. Once again, all parameters were fixed in an *ad hoc* fashion (here, they were fixed differently than before). Firstly, the parameter χ introduced proves itself fundamental in showing that two distinct regions regarding the average of gains and losses exist. High (low) values for this parameter are directly associated with low (high) volatility markets. Therefore, the introduction of such a parameter allows for the simulation of several different types of markets.

Also, due to the novel updating mechanism, it is found that the lattice evolves towards a self-organizing critical state (Bak, 1996). Self-organized criticality can be essentially defined as the tendency that certain systems have to organize themselves, purely due to the interactions amongst the constituents of the system (no external factors), towards a critical state. This critical state represents the system in some sort of unstable equilibrium, where minor perturbations may lead the system out of equilibrium. This event is known as an avalanche, and the distribution of the log frequency of avalanches per log step (by step we mean the smallest interval to go from one state of the system to another – it could be time, distance or other) follows a power law, exhibiting scale invariance. Another way of thinking about this phenomenon is by understanding that the laws which dictate dynamics are simple in the sense that they can be (most of the time) formulated analytically, and solutions can be computed, as long as these systems don't have many constituents. If the opposite is true, then the system becomes quite complex. Self-organized criticality is a mechanism which generates complexity and introduces catastrophism in the system, and it is suggested by Bak (1996) to be fundamental in understanding how nature works, even economic and financial nature (Bak et al., 1992).

In Dupoyet et al. (2012), the updating mechanism revolving around equation (5.77) drives the financial lattice into this critical state, by replacing its absolute maximum value with a random one extracted from (5.78) and (5.79). This could be interpreted as the market "correcting" itself every time financial returns, either positive or negative, are too high. Finally, the distribution of the log frequency of avalanches per log simulation time step follows a power law, thus substantiating the claim of observed self-organized criticality. In this situation, avalanches are interpreted as the set of updates during a single simulation time step.

By constructing gains distributions, it is observed the expected behaviour: higher-than Gaussian mean value and fat tails. Also, after simulating 100 lattice configurations and constructing their respective returns time series, it is assessed with the aid of a GARCH (1, 1) model, that these time series are, when compared with the NASDAQ historical data, indistinguishable, in terms of the parameters in both GARCH (1, 1) models.

Paolinelli & Arioli (2018) begins by considering the expression for the action (5.64) suggested by Ilinski (2001). Employing the change of variables, for the case of only two assets, cash and some hypothetical share, we have:

$$\psi_{i,t} = \sqrt{M\rho_{i,t}}\; e^{-\mathrm{i}\phi_{i,t}}$$
$$\bar{\psi}_{i,t} = \sqrt{M\rho_{i,t}}\; e^{\mathrm{i}\phi_{i,t}} \tag{5.82}$$

for $i = 0,1$, where $\rho_{i,t} \in [0,1]$ represents the density of money flows between assets, at time t, and $\phi_{i,t} \in [0,2\pi]$, where $\phi_{0,t} - \phi_{1,t}$ represents the "velocity" with which money flows. It is clear that, because of the assumption of a closed environment, we have $\rho_{0,t} + \rho_{1,t} = 1 \Longrightarrow \rho_t \equiv \rho_{0,t} = 1 - \rho_{1,t}$. Therefore, the perturbation (5.63) becomes:

$$F = 2\alpha\,(\rho_{t+1} - \rho_t) \tag{5.83}$$

and also the action itself must be changed to these variables. It allows for the execution of the numerical simulation.

Paolinelli & Arioli (2018) realize that a poor agreement with historical data is verified with these perturbations in the tail region. The relationship between the logarithms of the probability distribution of prices and the logarithm of the prices itself in this region motivates the following perturbation, suggested by the authors:

$$\sum_{k=1}^{J} 2\alpha_k\,(\rho_{t+1} - \rho_t)\,|\rho_{t+1} - \rho_t|^{\Gamma_k - 1} \tag{5.84}$$

where $J \geq 1$, $\Gamma_k \geq 1$ and α_k are integers. This allows for the possibility of sudden price changes, usually resultant of external factors of macroeconomic nature. When this happens, over- or under-estimations concerning the prices occur. These events are known as jumps. In fact, an interpretation for α_k and Γ_k is provided, which centres itself exactly on such notions. The term α_k, which is proportional to the volatility, has to do with probability of jumps occurring, while Γ_k influences the size of said jumps. The numerical simulations performed were all based on a Metropolis-Hastings algorithm, which is a Monte Carlo method. When comparing with historical data obtained from the S&P500 index and APPLE stocks[37], spanning three months, with a sampling rate around one minute, a good fit is observed when $J = 2$, even in the fat tails region, thus improving on Ilinski (2001) and Dupoyet et al. (2012) model. Also, the parameters responsible for the jumps are varied, one at a time, and the above interpretations are reinforced: increasing (decreasing) α_k, increases (decreases) the log distribution variance, thus increasing (decreasing) the probability of jumps, while increasing (decreasing) Γ_k, increases (decreases) the similarity with a lognormal distribution of the prices, thus decreasing (increasing) the size of the price jumps (for $k \in \{1,2\}$).

[37] Here, similar considerations regarding historical data to Dupoyet et al. (2010) and to Dupoyet et al. (2012) were employed, specifically, the usage of the log change in prices, per smallest time frame (one minute in this particular situation), as values from a stochastic variable.

To allow for different time scales, Paolinelli & Arioli (2019) introduced a different action. They began by noticing that the gauge-invariant quantities previously used, which are, in the continuous limit, essentially the following:

$$\begin{aligned} s_{Arbitrage} \equiv s_2 = \beta_2 \int_0^T (\partial_t \log S_t - r_t)^2 \, dt \\ s_{Gains} \equiv s_1 = \beta_1 \int_0^T |\partial_t \log S_t - r_t| \, dt \end{aligned} \tag{5.85}$$

are most appropriate for different time horizons. The first action is constituted by the risk-free gains possible by adopting an arbitrage trading strategy, while the second action is composed of the risky gains possible for some time horizon. Generally speaking, it's true that $s_1 > s_2$ for any time horizon. However, for a short time horizon, the value of the action associated with arbitrage is negligible when compared to the action of risky gains. For this reason, the following action is proposed:

$$s_p = \beta_p \int_0^T |\partial_t \log S_t - r_t|^p \, dt \tag{5.86}$$

where $p \in [1,\ 2]$, thus mixing both actions, depending on the chosen time horizon: p must be closer to one for short time horizons and it must be closer to two for longer time horizons. Now, regarding the probability density function, it must reflect the fact that experienced traders have greater chances of accruing profit over unexperienced ones, for the same levels of allowed arbitrage. This is accomplished by introducing a parameter $\lambda \in]0,\ 1[$, thus obtaining the probability density function:

$$P_\lambda(s_p) \propto \exp\{-\beta_p \cdot s_p^\lambda\} \tag{5.87}$$

The closer the parameter is to zero, less likely are small values for the action of happening, or conversely, greater amounts of returns are more probable in the short term, for the same amounts of arbitrage. Notice that for $p = 2$ and $\lambda = 1$, the "classical" distribution is obtained, in accordance with the geometric Brownian motion case (log-normality).

Similarly with the previous cases, the transition probability from S_0 to S_T is given by a path integral:

$$P(S_T|\, S_0) \propto \int e^{-\frac{1}{2\sigma^p} s_p^\lambda} D \log S_t \tag{5.88}$$

where $\beta_p = 1/(2\sigma^p)$, which is related with volatility.

Given that the action is not quadratic nor linear, there is no known analytical way of addressing this path integral. Therefore, the following approximation was suggested:

$$P\left(S_T \mid S_0\right) \approx$$

$$\approx \lim_{N\to\infty} \prod_{j=1}^{N-1} \int_0^{\infty} \exp\left\{-\frac{1}{2\sigma^p \Delta t}\left(\sum_{j=0}^{N-1} \left|\log S_{j+1} - \log S_j - r\right|^p \Delta t^{p-\frac{p^\lambda}{2^\lambda \lambda}}\right)^\lambda\right\} \frac{dS_j}{S_j} \tag{5.89}$$

which converges and has negligible relative errors. To carry on with the numerical simulation, the Quasi Monte Carlo method was employed for a ten dimensional integral, shown to be an adequate approximation.

To analyse in the short-term, the three month historical prices from three assets, AMAZON, GENERAL ELECTRIC and APPLE, were chosen, and three time frames were fixed: 1, 5 and 30 minutes. For each of this time frames, the parameters p, λ and σ were varied accordingly. The longer the time frame, the greater is the value for the first two parameters, ranging from 1.15 to 1.23 and from 0.15 to 0.23, respectively. For the long-term analysis, the thirty year historical prices from the Dow Jones and S&P500 indices were used[38]. The time frames were fixed to one day and one week, respectively. The parameters p and λ varied from 1.35 to 1.42 and from 0.35 to 0.42, respectively, and σ changed from 0.062 to 0.023.

It was found good agreement with historical data, for all time frames, specifically the one minute time frame, which, compared with the previously results, presents much better agreement. Also, p and λ were the same across all assets, which seems to suggest that they share similar dynamics. Furthermore, the variations experienced by the parameter σ are in agreement with its interpretation: share liquidity.

Finally, praise for this model is given not only for its agreement with empirical data, but also for the fact that it depends only on three parameters. Furthermore, it's non-Markovian by nature, separating it from the rest of the literature surrounding this subfield of finance.

5.2.4 Basic Framework Criticism

Some scepticism to the possibility of a gauge theory of finance was put forth early on, offering important criticism to the fundamental assumptions present in the theory.

Sornette (1998) targeted the beginnings of this theory, which were introduced in subsection 5.2. The author presents, essentially, three important observations:

[38] The choice for indices instead of assets was justified by noticing that across so many years, asset prices may be affected by other external factors, which would introduce further complications.

1. There is no explanation for the use of probability weights that follow a Boltzmann distribution, $P\left(\left\{U_{ij}\right\}\right) \propto e^{-\beta \cdot s}$, seen as an *ad hoc* choice meant to resemble with the physical theories which served as inspiration;

2. It is suggested that the existence of uncertainty, formalized as noise, is responsible for the introduction of "virtual" arbitrage opportunities, but in the case of a complete market, random variables have the Markov property, implying the nonexistence of arbitrage opportunities[39], despite the fact that noise is obviously present;

3. There's no reason to believe that elementary arbitrage strategies serve as building blocks for other, more complex, strategies, even though this is tacitly implied in the theory.

With respect to the first observation, a possible justification might rest on the fact that, in principle, any trader in the market can follow an arbitrage strategy, allowing for the possibility of generating risk-free profit in a consistent fashion. Therefore, big arbitrage opportunities must be less probable than smaller ones, thus motivating the choice for such probability (Paolinelli & Arioli, 2019)

The other two points remain unaddressed. Regarding the third one, there's even an example of a different arbitrage strategy which doesn't follow the elementary *plaquette* construction. This example was suggested by Young (1999), where an arbitrage strategy involving a single asset at three different times can be exploited by making use of the different interest rates.

A fourth one is also presented, which concerns the log-normality derived using this formalism, seen as a verification of the theory's validity, since it reproduces the "classical" result. Sornette (1998) comments on this by stating that correspondence between any new theory and its old equivalent in some "classical" setting does not constitute substantial proof.

Further criticism was presented by Sokolov et al. (2010), regarding, in particular, how the transition probability (5.58) was derived. When constructing the action in the continuous time limit, that is, when $\Delta t \to 0$, it is assumed that $a_j \Delta t \to 1/\left(2\sigma^2\right)$, with no justification.

However, when deriving an important component of the transition probability (known in physics as the Hamiltonian, which can be thought of as an operator responsible for dictating how some quantity changes throughout time), Ilinski (2001) arrives to a quotient where in the denominator is Δt, meaning one of two things: either $\Delta t \to 0$

[39] This follows from the fundamental theorem of finance, where it is proven that if there exists a probability measure such that the random variable is a martingale, then the market is arbitrage-free.

and the Hamiltonian tends to infinity or there exists some inconsistency in the theory. Furthermore, the transition probability (5.58) has a determinant equal to zero, implying that it can't correctly identify an evolution operator.

Also, regarding the quantity being integrated in the action (5.64), analogous to what is known in physics as the Lagrangian, Sokolov et al. (2010) noticed an algebraic mistake in the derivation of the "equations of motion" for financial markets, obtained by minimizing the Lagrangian. After correcting for such a mistake, solutions for these differential equations are unstable and present impossible values for interest and exchange rates. Since these are derived from the Lagrangian, it means that some incorrect detail is present in it.

These criticisms were not, in any way, addressed by the authors of the numerical simulations, or theoretical extensions. The fact that there exists this inconsistency when one goes from the discrete to the continuous case was not addressed in the previous section, but since the numerical simulations are all performed in a discrete environment, this doesn't really constitute a problem. However, the fact that the transition probability is wrongly constructed is a problem which also affects the applications previously exposed.

Nevertheless, good agreement with historical data is observed, implying, at least partly, that the descriptive power of the formalism is still present, despite its many inconsistencies. Still, in order for the model to become competent in describing market reality it must be built in solid theoretical grounds, otherwise, the times where it "gets" reality right, are of no use, since there is no self-consistent framework, derived from first-principles.

5.3 Deflator-Term Structure Gauge Symmetry [40]

In the attempt to provide a geometrical framework to stochastic finance, Farinelli (2015) introduced a number of geometrical reinterpretations of financial phenomena, stochastic in nature. To accomplish this, the author utilized most of the ideas shown in 5.2, but in a stochastic way, with some important distinctions, such as a different total space "hovering" above the base space and a different gauge. The latter was greatly inspired by Smith & Speed (1998), where notions of gauge theory are implicitly present (Farinelli, 2015).

As stated above, most of the formalism to be presented follows from 5.2, but in stochastic terms. For this reason, let there be a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, where Ω is the

[40] Later in section 5.3, it will be shown that other symmetries were explored, besides gauge symmetries. Also, the model developed by Farinelli (2015) encapsulates the one shown in 5.2, and therefore, the notion of *numéraire* invariance is also present.

space of all possible events, the filtration $\mathscr{F}$ is the set of all possible subsets of Ω (a $\sigma-$algebra), essentially denoting the available market information at each time t, and $\mathbb{P}$ is the probability of each event happening (see, for example, Hunt & Kennedy (2004) or Shreve (2004) for an exposition of stochastic finance). Randomness is therefore modelled via this formalism (as is standard in financial literature). All quantities (e.g. prices, interest rates) to be considered below are assumed to be stochastic processes: for each $t \in T$ and possible event $\omega \in \mathscr{F}_t$, the considered quantity assumes some value. Also, these stochastic processes are assumed to be semimartingales (any stochastic process that can be decomposed into a local martingale – a term whose expected value is independent of past events – and a "drift" term), where its dynamics are given by Brownian motions coupled with a drift term, similar to equation (5.37) in section 5.2, unless otherwise stated. In what follows, transaction costs are non-existent, short sales are allowed and economic agents have access to the same information (represented by the filtration).

Let a financial market be characterized by a finite amount of assets $I = \{i|\, i = 0,\ 1,\ 2,\ ...,\ N\}$ traded in continuous time $T = \{t|\, t \in [0, +\infty[\}$. Each asset i is nominally priced as S_t^i, where for the 0^{th} asset, interpreted as before as cash in a bank account, it is given by the strictly positive $S_t^0 = \exp\left\{\int_0^t r_\tau^0 d\tau\right\}$, where r_t^0 represents the continuous interest rate[41]. From an investor perspective, an investing strategy is characterized simply by the amount of each asset that the investor holds at some time $t \in T$, denoted by x_t^i. A portfolio is then a vector $x_t = \left(x_t^0,\ x_t^1,\ ...,\ x_t^N\right)$ in $\mathbb{R}^{N+1}$, whose value is then:

$$V_t^x = \sum_{i=0}^{N} x_t^i S_t^i = x_t \cdot S_t \tag{5.90}$$

Opposing the formalism presented in 5.2, the rationale of Smith & Speed (1998) was followed. It stems from the realization that each asset can be fully identified by its value at time t, measured in terms of some *numéraire*, and by its expected forecasted values at that time, in terms of the asset's value. For this reason, each asset i will be mapped to the pair $\left(D_t^i,\ P_{t,s}^i\right)$, where D_t^i is denoted as the deflator and $P_{t,s}^i$ is denoted as term structure. It's with pairs of deflators and term structures, denoted in Farinelli (2015) and Smith & Speed (1998) as gauges[42], that financial instruments can be modelled.

Deflators are the values of each asset relative to some *numéraire*. Notice that it doesn't have to be the cash account. In fact, the only prerequisite for the choice of a *numéraire* is that it must be an asset whose price is a strictly positive stochastic process. Some portfolio

[41] This particular semimartingale is said to be predictable since its value is known in advance at each t, even though it changes throughout time.

[42] Calling it gauges cements the idea that each asset is measured in this way.

of assets could be a *numéraire*, if this condition is met[43]. Therefore, given a price vector $S_t = \left(S_t^0,\, S_t^1,\, ...,\, S_t^N\right)$ and portfolio weight $b_t = \left(b_t^0,\, b_t^1,\, ...,\, b_t^N\right)$ whose value is strictly positive, we have:

$$D_t = \left(\frac{S_t^0}{b_t \cdot S_t},\, \frac{S_t^1}{b_t \cdot S_t},\, ...,\, \frac{S_t^N}{b_t \cdot S_t}\right) \tag{5.91}$$

If $b_t = (1,\, 0,\, ...,\, 0)$, then the portfolio used as *numéraire* is nothing but the cash asset, and thus we obtain:

$$D_t = \left(1,\, \hat{S}_t^{\,1},\, ...,\, \hat{S}_t^{\,N}\right) \tag{5.92}$$

where $\hat{S}_t^{\,i}$ represents the price of asset i in terms of current cash units. In other words, it compares the value of the asset with the value of money. Notice that ratios between deflators give relative prices (traded amounts in a barter economy), and if the asset in consideration is a currency, deflators represent the time evolution of inflation, and ratios denote exchange rates.

Term structures are values at time t of a synthetic zero-coupon bond which delivers at time s a unit of asset i (similar to a forward contract, which delivers the asset at some future time, at a price established presently). Essentially, it represents the possible forecasts of each asset's deflator value, with respect to the chosen *numéraire*: how much would an investor be willing to pay right now to receive asset i at some future time s, for all $s \geq t$. It's assumed that $P_{t,t}^i = 1$, that is, the price of a bond delivering asset i right now is the price of the asset itself, $P_{t,s}^i > 0$ and $\lim_{s\to\infty} P_{t,s}^i = 0$ exponentially fast. If $f_{t,s}^i$ is the instantaneous forward rate, representing the forecasted future discount rate, then:

$$P_{t,s}^i = \exp\left\{-\int_t^s f_{t,\tau}^i d\tau\right\} \tag{5.93}$$

where $r_t^i = \lim_{\Delta t\to 0} f_{t,t+\Delta t}^i$ is the short rate[44]. With this in mind, the product $D_t^i P_{t,s}^i$ is interpreted as the amount one would pay at time t to receive D_t^i at some future time s (Smith & Speed, 1998). Notice that as s tends to infinity, the amount one is willing to pay to receive D_s^i is practically zero, reflecting the fact that greater uncertainty is present for long time horizons.

This way, a more general version of the market is presented, where each of the existent assets is modelled by a gauge: a pair of deflator and term structure. Once again, it's

[43] This idea of indifference towards how prices are measured served as the gauge symmetry in section 5.2.

[44] If $f_{t,s}^i > 0$ for all $t, s \in T$, then it's said that the term structure satisfies the positive interest condition. This is appropriate if the assets one considers are more valuable in the present than in the future (storable financial instruments e.g. non-perishable goods). If the instantaneous forward rate doesn't follow such a condition, the pair of deflator and term structure (the gauge) is said to be a principal gauge.

important to understand that the choice of the gauge is not unique, since the set of deflators can always be multiplied by some positive semimartingale, thus returning a new set of deflators which still characterize the market, as long as the term structure is changed appropriately. This was the invariance exploited in 5.2.

However, because each gauge describes assets via a zero-coupon bond, any kind of future cashflows are not being considered e.g. coupons or dividends for the case of a coupon bond or some generic share, respectively, which clearly affects the asset's value and potential forecasts. Therefore, a gauge transform $(D,\, P) \mapsto g((D,\, P) = ((D^g,\, P^g)$ is defined as follows:

$$D_t^{i,g} = D_t^i \int_0^{+\infty} g_\tau^i P_{t,t+\tau}^i d\tau \tag{5.94}$$

$$P_{t,s}^{i,g} = \frac{\int_0^{+\infty} g_\tau^i P_{t,s+\tau}^i d\tau}{\int_0^{+\infty} g_\tau^i P_{t,t+\tau}^i d\tau} \tag{5.95}$$

where g_t^i denotes a stochastic cashflow, associated with asset i at time t, determined as:

$$g_t^i = -\mathop{\mathbb{E}}_t \left[\frac{d}{dt} D_t^i \right] + r_t^0 D_t^i \tag{5.96}$$

where r_t^0 denotes the continuous interest rate of the *numéraire* [45]. These cashflows[46] are measured in terms of the deflators. One can interpret equation (5.94) as the value of an asset i that yields additional cashflows, in terms of some previously defined *numéraire*. In fact, the integral in (5.94) is nothing but the discounted cashflows, summed throughout all future times. The new term structure (5.95) is defined posterior to the deflator. Multiplying both equations (5.94) and (5.95), a suggestive expression appears:

$$D_t^i P_{t,s}^i \xrightarrow{g} D_t^{i,g} P_{t,s}^{i,g} = \int_0^{+\infty} g_\tau^i \left(D_t^i P_{t,t+\tau}^i \right) d\tau \tag{5.97}$$

It represents the amount one is willing to pay at time t to receive $D_t^{i,g}$ at time s, taking into account all future cashflows that the investor will receive for holding the asset.

With the above in mind, the market is modelled as a fibre bundle, similarly to the previous sections. For such purpose, from the $N+1$ assets, one is fixed as the *numéraire*. As base space, the manifold considered was the set of all possible portfolio holdings at each time:

[45] The expected time derivative in (5.96) is the stochastic derivative, from now on denoted as $\tilde{\partial}_t$. Technically speaking, it stands for the derivative associated with the Stratonovich integral, instead of Itô's, thus satisfying the chain rule, and permitting a differential geometric treatment (Farinelli, 2015).

[46] For the gauge transform to be well defined, it is necessary that:

$$\lim_{t \to +\infty} \sup \exp \left(\frac{\log |g_t^i|}{t} \right) \leq 1$$

$$M = \left\{(x,t) \,|\, x \in \mathbb{R}^{N+1}, t \in T\right\} \tag{5.98}$$

The gauges for each possible portfolio are as follows:

$$D_t^x = \sum_{i=1}^{N} x_i D_t^i \tag{5.99}$$

$$P_{t,s}^x = \exp\left\{-\frac{1}{D_t^x} \sum_{i=1}^{N} x_i D_t^i \int_t^s f_{t,\tau}^i d\tau\right\} \tag{5.100}$$

Clearly, (5.99) follows immediately from (5.90) and (5.100) can be thought of as a weighted average of the instantaneous forwards rate associated with the portfolio. The total space E, denoted as the market fibre bundle, is thus defined as:

$$M = \left\{\left(D_t^{x,g}, P_{t,s}^{x,g}\right) \Big| \, (x,t) \in E, \; s \in T, \; g \in G\right\} \tag{5.101}$$

Employing the machinery introduced in Chapter 2, section 2.3, there exists a projection map between the total and base space:

$$\begin{aligned} \pi : E &\longrightarrow M \\ \left(D_t^{x,g}, P_{t,s}^{x,g}\right) &\longmapsto (x,t) \end{aligned} \tag{5.102}$$

which represents the idea that each portfolio has to it an associated gauge, corresponding to that portfolio's value, with a specific cashflow structure g. The set of all possible cashflow profiles denotes the fibre, which coincides with the structural group [47]. Therefore, the fibre bundle is a principal one. This group acts on the total space similarly to (2.7):

$$\begin{aligned} E \times G &\longrightarrow E \\ \left(D_t^x, P_{t,s}^x, \; g\right) &\longmapsto \left(D_t^{x,g}, P_{t,s}^{x,g}\right) \end{aligned} \tag{5.103}$$

Finally, we recall that there exists local trivializations for the total space, shown in (2.6). However, because the total space at hand is in fact globally trivial, we have $E = M \times G$, given by:

[47] Notice that the group action is the convolution operation:

$$(g * h)_t = \int_0^t g_\tau h_{t-\tau} d\tau$$

For any $g, h \in G$. Also, not every cashflow is part of group G. The prerequisite is that for every element of $g \in G$ there exists $q \in G$ such that $(g * q)_t$ is equal to the Dirac function (zero everywhere and infinite at the origin).

$$\left(D_t^{x,g}, P_{t,s}^{x,g}\right) = (x,t,g) \tag{5.104}$$

In other words, a portfolio x at time t with a cashflow structure g is identified by its deflator $D_t^{x,g}$ and term structure $P_{t,s}^{x,g}$, representing the portfolio's value. This fibre bundle construction is a more generalized version of the one exposed in 5.2.

Now, consider a path in the base space, connecting the points $(x_0,\ t_0)$ and $(x_1,\ t_1)$, where the indices denote the initial and final times of the investment (in other words, the index 0 denotes the time when the investor acquired the portfolio x, while the market has been "operational" for t time). A section of the fibre bundle corresponding to the point $(x_0,\ t_0)$ is an associated cashflow structure g_0, which is an element of the fibre. Applying the parallel transport operator, one gets the cashflow structure g_1 associated with the fibre of $(x_1,\ t_1)$. The connection 1-form chosen by the author was:

$$A(x,t,g) = \frac{1}{D_t^x}\sum_{i=1}^{N} D_t^i dx_i - r_t^x dt \tag{5.105}$$

where $r_t^x = \lim_{s\to t} \frac{\partial}{\partial s}\log P_{t,s}$. From now on, the term structure will be identified with the continuous interest rate of the portfolio. Recalling the covariant derivative equation (2.25), then:

$$\begin{aligned}
&\left(\tilde{\partial}_t - A(x,t,g)\right) g = 0\\
\implies g_1 = &\begin{cases} g_0 \frac{D_t^{x_1}}{D_t^{x_0}}, & \text{if the path is in the portfolio direction}\\ \\ g_0 \exp\left\{-\int_0^1 r_\tau^{x_\tau} d\tau\right\}, & \text{if the path is in the time direction}\end{cases}
\end{aligned} \tag{5.106}$$

thus encapsulating the notion of a connection as prices, exchange rates (the quotient of deflators) and interest rates (the integral of the term structure), as in 5.2.

Curvature is also calculated, using equation (2.29):

$$R(x,t,g) = \sum_{i=1}^{N} \frac{\partial}{\partial x_i}\left(\tilde{\partial}_t \log D_t^x + r_t^x\right) dt \wedge dx_i \tag{5.107}$$

where $\wedge$ denotes the wedge product, which is essentially a cross product, but generalized for any dimension. Therefore, one can see that the fibre bundle is flat if:

$$\tilde{\partial}_t \log D_t^x + r_t^x = \text{Const}_t \tag{5.108}$$

where Const_t is a constant with respect to portfolio nominals. If we let this constant be a strictly positive semimartingale β_t, then we have:

$$\tilde{\partial}_t \log(\beta_t D_t^x) + r_t^x = 0 \tag{5.109}$$

this constant is called in the financial literature as the state price deflator or pricing kernel, and its existence is a consequence of a market which doesn't allow for arbitrage opportunities[48]. Therefore, there exists a link between no-arbitrage and no curvature, consistent with what was previously shown in 5.2. A financial interpretation of (5.109) is that a market possesses no risk-free opportunities if and only if changes to the log value of any submarket (any set of portfolios) is solely caused by the flow of short rate through the boundary of the submarket (some portfolio which bounds the rest). This constitutes a continuity equation, which holds only if the market environment is arbitrage free.

Farinelli (2015) then proceeds to establish similar axioms to those presented in 5.2.3.1, but in a stochastic context, and with a different fibre bundle, as shown before. An arbitrage action is built, which must be minimized. For that end, the following action is considered:

$$s^{\beta}(\gamma;\, D, r) = \int_{\gamma} \left\{ \tilde{\partial}_t \log D_t^x + r_t^x \right\} ds \tag{5.110}$$

where β_t is any positive semimartingale and $ds = |\tilde{\partial}_t x| dt$ is an infinitesimal step in the portfolio-time space. Let the trading path γ be self-financing[49] and be the stochastic discount factor. Then (5.109) can be rewritten, and the following inequality holds if there exists arbitrage opportunities in the market:

$$s^{\beta}(\gamma;\, D, r) \neq 0 \Longrightarrow \frac{\beta_1 D_1^{x_1}}{d_{0,1}^{\gamma}} \neq \beta_0 D_0^{x_0} \tag{5.111}$$

Furthermore, the author explores the relationship between self-financing closed strategies in the base space and the existence and effects of associated arbitrage opportunities. It is found that if this kind of closed strategies can be smoothly deformed to a single point, then arbitrage is non-existent. The intuition behind this is that, in a market where it is impossible to gain profit without risk, trading some portfolio in such a

[48] This is known as the Fundamental Theorem of Asset Pricing, which basically states that the market allows no risk-free opportunities if an only if there exists an equivalent probability measure such that the price processes are martingales, which in turn means that there must exist pricing kernel such that the discounted price process in terms of the pricing kernel is a martingale (Hunt & Kennedy, 2004).

[49] That is, any changes in value of the portfolio are solely due to changes in the price of the price deflators, not on the nominal weights. However, since we are using the Stratonovich's derivative, we have:

$$\left(D_t^x \cdot \tilde{\partial}_t x \right) = -\langle x, D \rangle_t / 2$$

where $\langle \cdot, \cdot \rangle_t$ denotes the quadratic covariation between two stochastic processes, for all $x \in \gamma$.

way that the beginning and ending of the trade are the same portfolio, or holding that portfolio throughout the investment time horizon, amounts to exactly the same gains – no arbitrage gains are acquired. Therefore, it's this kind of "obstruction" in the portfolio nominal space that allows arbitrage opportunities. Also, Farinelli (2015) points out that instead of considering every possible trading path, one can consider the set of all closed, self-financing trading paths which can be smoothly transformed to one another, by simply varying the portfolio nominals. It is found that the arbitrage action $s^{\beta}(\gamma;\, D, r)$ and associated stochastic discount rate $d^{\gamma}_{0,1}$ are the same for each possible set.

Finally, the market dynamics are considered. These are a result of the feedback between market portfolio strategies and the deflators and short rates dynamics, happening simultaneously. In order to achieve this, the arbitrage action must be minimized. With this in mind, it must be written as the integral of the Lagrangian which is minimized (see 2.2.2) by stating that infinitesimal variations to its expected value, resultant from infinitesimal variations of points in the total space (that is, changes in the path of portfolio nominals , and associated deflators and short rates), must be zero. If the Lagrangian is:

$$L(q_t, \dot{q}_t) = |\tilde{\partial}_t \tilde{x}| \frac{x_t \cdot \left(\tilde{\partial}_t D_t + r_t D_t\right)}{x_t \cdot D_t} \tag{5.112}$$

where $q_t \equiv (x_t, D_t, r_t) \in E$ and $\dot{q}_t \equiv \left(\tilde{\partial}_t x_t, \tilde{\partial}_t D_t, \tilde{\partial}_t r_t\right)$, which follows immediately from (5.110), then the arbitrage action is minimized if the following holds:

$$\tilde{\partial}_t \frac{\partial L(q_t, \dot{q}_t)}{\partial \dot{q}} - \frac{\partial L(q_t, \dot{q}_t)}{\partial q} = 0 \tag{5.113}$$

under the self-financing portfolio constraint (see footnote 49). This equation is known as the Euler-Lagrange equation, but in a stochastic context. The author calculates equation (5.113) by assuming that q_t is equal to a deterministic component $\mathbb{E}_0[q_t]$ and by a perturbation δq_t of zero mean, and that various stochastic conditions must be satisfied (see Farinelli (2015) for a description of such conditions). It is concluded that the tuple (x_t, D_t, r_t) which minimizes arbitrage (that is, which are solutions to (5.113)) must be of a certain form. The market portfolio must be:

$$x_t = x_0 + \delta x_t \tag{5.114}$$

denoting that the portfolio at time t is given by the initial market portfolio added a perturbation that varies with time. The deflator must be:

$$D_t = e^{-t} g(x_0, D_0) + \delta D_t \tag{5.115}$$

where $g(x_0, D_0)$ is an operator essentially representing the value of all possible combinations of assets in the market which can't be represented in the initial market portfolio, in terms of its value, and the expected future value of the initial market portfolio. This way, the deflator which characterizes the value of the market portfolio (in terms of some *numéraire*) changes throughout time either by changes of its value or by the exclusion or inclusion of assets in the market portfolio. Finally, the short rate is given by a complicated expression of exponential integrals with terms similar to the right hand side of (5.115), excepting δD_t.

Farinelli (2015) points out that equation (5.113) is invariant under gauge transformations similar in form with (5.103). Also, the Lagrangian in (5.112) exhibits symmetry with respect to translations of the portfolio nominals and deflators (expressed by the multiplication with any possible $\mathrm{e} \in \mathbb{R}^N$). By Noether's theorem (in a stochastic version, see Farinelli (2015)), symmetries imply conservation laws, which lead to the following system of equations that the dynamics of market portfolios and deflators must satisfy, for a self-financing strategy:

$$\begin{cases} \frac{d}{dt}\mathbb{E}_0\left[\frac{\mathrm{e}\cdot x_t}{x_t\cdot D_t}\right] = 0 \\ \\ \frac{d}{dt}\mathbb{E}_0\left[\frac{x_t\cdot\left(\tilde{\partial}_t D_t + r_t D_t\right)}{x_t\cdot D_t}\frac{\mathrm{e}\cdot\tilde{\partial}_t\tilde{x}_t}{\left|\tilde{\partial}_t\tilde{x}_t\right|}\right] = 0 \end{cases} \tag{5.116}$$

Farinelli (2015) concludes by noting that for the case of no-arbitrage, the system of equations above returns the classical result from stochastic finance, stating that the deflator stochastic process, which, in terms of cash *numéraire*, denotes prices, is a martingale under a possible change of probability measure to a risk neutral one – the Fundamental Theorem of Asset Pricing (see footnote 48).

6 CONCLUSIONS

In this dissertation, applications of gauge theory to finance were explored with the recourse of a systematic literature review. After an initial search, it was found that current applications occupy the subfield of financial markets, regarding topics such as asset pricing, option pricing and general market dynamics. The pool of pertinent articles was composed by 14 papers, all published in physics' journals. The usage of the gauge theoretical formalism can be differentiated into three approaches, under the criteria of the different gauge symmetries exploited:

- 1^{st} approach: changes in the beliefs and preferences of investors leave the differential equation dictating the option's value through time invariant, after some modifications transforming it into a covariant equation.
- 2^{nd} approach: changes in the price units (a change of *numéraire* - that with which prices are measured, usually cash in some currency), and prices itself, don't affect some quantities, which are postulated to be the ones with which market dynamics can be estimated;
- 3^{rd} approach: portfolios of assets can be solely described by its prices throughout time, by its expected forecasts and its cashflows. Changes to the latter represent changes to the financial instrument's characteristics, but preserve its nature (for example, the difference between a portfolio of a single bond and a portfolio of a single coupon bond is the cashflows that one accrues, but the financial instrument itself is the same). Similarly to the 2^{nd} approach, it's assumed that the dynamics are invariant under these kind of changes.

The 2^{nd} approach motivated further development and criticism. A generalization for any size portfolios, instead of a single asset (Morisawa, 2009), and an application to option pricing, employing a similar rationale as in the 1^{st} approach, where a covariant equation is established (Zhou & Xiao, 2010), were proposed. We note that the latter could be combined with the gauge-symmetry encountered in the 1^{st} approach, in order to achieve a covariant option pricing equation which takes into account two kinds of gauge-symmetry. However, there exists some differences between the fibre bundle constructions from both, and modifications are due. Furthermore, numerical simulations of the 2^{nd} approach were performed, mostly concerning small time horizons (high frequency trading), and good agreement was obtained when compared with historical data (Dupoyet et al., 2010, 2012; Paolinelli & Arioli, 2018, 2019).

These simulations correctly derive the characteristic shape of the prices probability distribution, exhibiting the expected fat tails and higher-than-Gaussian probability around the mean value. These stemmed from the notion of optimizing a functional known in physics' literature as the action, constituted by the gains from arbitrage opportunities and from wealth-maximizing investment strategies, denoted as money flows (Ilinski, 2000; Paolinelli & Arioli, 2018). Afterwards, inspired by quantum field theory formalism, each action functional was used as an exponential probabilistic weight, representing the most probable price and portfolio (of a single asset and cash in the case of Dupoyet et al. (2012) and Paolinelli & Arioli (2018)) trajectories in the market: high arbitrage is less probable and investment strategies with high expected gains are more probable (thus reflecting the wealth-maximizing behaviour of the investors participating in the market, when aware of some level of information).

Building upon this, other aspects were introduced, such as probabilistic weights dependent of the investment time horizon, which are scale invariant, thus reproducing the effect of various assets' holding times (Dupoyet et al., 2010), an updating strategy which drives the whole system into a state of self-organized criticality (Dupoyet et al., 2012), perturbations replicating price jumps (Paolinelli & Arioli, 2018) and a different action whole together, appropriate for different time horizons (Paolinelli & Arioli, 2019). For further research, it is suggested the investigation of how could certain perturbations be introduced in the model (e.g. inflation rates or dividend growth) that break the gauge symmetry, thus becoming closer to a real-life scenario, the self-organized criticality that can be achieved, the modelling of sudden price jumps, and the exploration of the parameters who mediate different time horizons, especially longer time periods. Overall, the notion of gauge invariance (at least regarding the local *numéraire* gauge symmetry) seemed to be realized, and hopes were expressed for the capability of modelling the probability distribution which dictates future prices.

Still, the criticism put forth by Sornette (1998) and Sokolov et al. (2010) wasn't completely, or not at all, addressed. One could argue that the probabilistic weight for the action being of exponential nature mirrors the fact that the most probable market state is for it to not allow arbitrage opportunities, hence the strictly decaying function, but a more profound reason is lacking, as pointed out by Sornette (1998). What was criticized by Sokolov et al. (2010) remains valid, and unaddressed, despite the fact that the models built by Dupoyet et al. (2010; 2012) deviate slightly from the problems here suggested, and still produce good results. In fact, as already mentioned, good results were achieved, despite the hovering mistakes. This implies the need for further research. Regarding the borderline arbitrarily chosen arbitrage strategies, the 3^{rd} approach, which can be understood, to some extent, as a more generalized version of the 2^{nd} approach's work, touches briefly on this, and attempts to generalize the kind of trading paths which deliver arbitrage opportunities. This could motivate further work exploring which trading paths can actually be called elementary, spanning others, more complicated ones. Also, different symmetries are explored, and benefits to risk management, pricing of assets and detection of statistical arbitrage are mentioned.

Unfortunately, no further published papers tackled what was introduced in the 3^{rd} approach, from a simulation standpoint, thus lacking any possibility to substantiate these claims. Also, we would like to comment on one of the main advantages, which is the capacity to describe reality, aided by models, derived from first principles, with little to no *ad hoc* assumptions. For example, when one compares with other known models in finance, many possess the descriptive power, but are tailored to a particular phenomenon (Dupoyet et al., 2010). The theoretical details here exposed don't suffer from the same problem. However, we note that the models all required some sort of additional perturbation, which could be argued to be an *ad hoc* addition, on hopes of accounting for some phenomena. More work regarding the perturbations to be used, and how to anchor them with solid theory, should be done.

In the process of understanding how gauge theories have been applied to finance, a reconceptualization of financial markets as geometrical objects was necessary, and it proved itself quite successful in providing known financial entities a clear geometrical interpretation. Probably the most interesting one was the relationship between arbitrage and curvature, which, as it was shown, relates to gauge-invariant quantities with which the dynamics can be described. The notion of arbitrage is fundamental in financial literature, allowing for the unambiguously pricing of any financial instrument (at least in theory) by assuming that any investment has always associated to it some risk. Examples of this no-arbitrage argument appear and reappear constantly throughout the literature. Assuming the existence of arbitrage opportunities, at least temporarily and sporadically, one could be closer to be able to describe non-equilibrium phenomena (Smolin, 2009), which in reality exist, and can affect markets in very consequential ways. Instead of equilibrium, one would observe steady-state equilibrium, and non-equilibrium phenomena would deviate the markets outside of equilibrium in a fluctuating manner, and then the corrective market forces would bring it back to an equilibrium scenario (efficient market hypothesis).

Another interesting aspect of the 2^{nd} approach's gauge symmetry, that can be further investigated, is its importance in non-equilibrium asset pricing. As we've seen, the idea that prices can be transformed via a global gauge transformation, the change of measuring units – *numéraire* – which affects all investors, has important implications in modelling market dynamics. However, in an out-of-equilibrium market, this symmetry has the potential of being more powerful. In this scenario, some assets may not have a clearly defined price, but since they're being traded, some price must be attached to them. These prices are essentially arbitrary, and the gauge-symmetry above mentioned can now be understood locally instead of globally: each investor is free to rescale the price of any asset or portfolio owned. Under this gauge transformation, the exchange rates between any other asset change appropriately. What is speculated is that even if the share's value changes due to the sheer convictions of the investor holding it, the gains associated with arbitrage strategies (and other possible gauge invariant quantities) remain the same. From this observation, the notion of prices as purely relative quantities is manifested, and the dynamics of the market can be considered outside of equilibrium (Smolin, 2009). In fact, an out-of-equilibrium case was also considered by Farinelli (2015), utilizing an operator which resembles an idea introduced by Malaney (1996) in her PhD thesis: the notion that changes to the assets' price in an out-of-equilibrium situation, due to the arbitrary nature of prices in this case, must be mediated by barter. This becomes quite intuitive if one considers the case of new assets being introduced in the market. Exactly because of their novelty, there is no way of comparing them, value-wise, with pre-existent assets, and for that reason, the price attached to this new asset is inevitably a mismatch. Thus, we return to the idea of local price changes as gauge transformations, as long as we are dealing with a non-equilibrium situation.

Finally, we comment on the chosen methodology. This systematic literature review allowed us to, in a consistent and replicable manner, investigate the contributions onto a new topic, filled with potential. Instead of choosing articles for review in an arbitrary, or at best, subjective way, as one usually sees with traditional literature reviews, with a systematic literature review one strives for objectivity, at least to some extent. By setting a group of keywords assumed to be of enough descriptiveness regarding the field we were interested with, and introducing them in a vast database, one should expect all articles described by such keywords to show up. The next step, of exclusion and inclusion, is inevitably riddled with subjectivity, which we tried to minimize by exposing the best we could each step of the way, with constant advising from the experts who aided the development of the chosen set of keywords.

Lastly, we would like to point out that, given the relative infancy of this subject, many contributions are found not in the form of articles, but in working papers. For further research on this subject, we would advise to complement with these documents. Also, regarding the systematic literature review itself, a different choice of keywords could lead to other results. However, due to its nature, we are confident about the acquired pool of articles, thus reflecting the strength of the methodology.

REFERENCES

Baaquie, B. (2004) *Quantum Finance: Path Integrals and Hamiltonians for Options and Interest Rates*, Cambridge University Press.

Bak, P., Chen, K., Scheinkman, J.A. & Woodford, M. (1992) *Self-organized Criticality and Fluctuations in Economics*, Working Paper.

Bak, P. (1996) *How Nature Works: the Science of Self-Organizing Criticality*, Springer-Verlag New York.

Cochrane, A.L. (1999) *Effectiveness and Efficiency: Random Reflections on Health Services*, London, Royal Society of Medicine Press Ltd.

Denyer, D., & Tranfield, D. (2009) Producing a Systematic Review. In *The Sage Handbook of Organizational Research Methods*. (Ch. 39, pp. 671-689) London, Sage Publications Ltd.

Dupoyet, B., Fiebig, H.R. & Musgrove, D.P. (2010) Gauge Invariant Lattice Quantum Field Theory: Implications for Statistical Properties of High Frequency Financial Markets. *Physica A: Statistical Mechanics and its Applications*, 389 (1), 107-116.

Dupoyet, B., Fiebig, H.R. & Musgrove, D.P. (2011) Replicating financial market dynamics with a simple self-organized critical lattice model. *Physica A: Statistical Mechanics and its Applications*, 390 (18-19), 3120-3135.

Dupoyet, B., Fiebig, H.R. & Musgrove, D.P. (2012) Arbitrage-Free Self-Organizing Markets with GARCH Properties: Generating them in the Lab with a Lattice Model. *Physica A: Statistical Mechanics and its Applications*, 391 (18), 4350-4363.

Durach, C.F., Kembro J. & Wieland, A. (2017) A New Paradigm for Systematic Literature Reviews in Supply Chain Management. *Journal of Supply Chain Management*, 53 (4), 67-85.

Farinelli, S. (2015) Geometric Arbitrage Theory and Market Dynamics. *Journal of Geometric Mechanics*, 7(4), 431-471.

Feynman, R.P. (1948) Space-Time Approach to Non-Relativistic Quantum Mechanics, *Reviews Of Modern Physics*, 20 (2), 367-387.

François, J., Lazzarini, S. & Masson, T. (2014) *Gauge field theories: various mathematical approaches*. Working Paper, Université de Toulon.

Hunt, P. & Kennedy, J. (2004) *Financial Derivatives in Theory and Practice*, Revised Edition, Wiley.

Ilinski, K. (1997) *Physics of Finance*, Working Paper, University of Birmingham.

Ilinski, K. (2000) Gauge Geometry of Financial Markets. *Journal of Physics A: Mathematical and Theoretical*, 33 (1), L5-L14.

Ilinski, K. (2001) *Physics of Finance: Gauge-Modelling in Non-equilibrium Pricing*, John Wiley & Sons.

Isham, C.J. (1999) *Modern Differential Geometry for Physicists*, 2^{nd} Edition, World Scientific Lecture Notes in Physics - Vol. 61.

Kholodnyi, V. (2002a) Valuation and Dynamic Replication of Contingent Claims in a General Market Environment based on the Beliefs-Preferences Gauge Symmetry. *Russian Journal of Mathematical Physics*, 9 (2), 161-179.

Kholodnyi, V. (2002b) Valuation and Dynamic Replication of Contingent Claims in the Framework of the Beliefs-Preferences Gauge Symmetry. *European Physical Journal B*, 27 (2), 229-238.

Kholodnyi, V. (2003) Beliefs-Preferences Gauge Symmetry and Dynamic Replication of Contingent Claims in a General Market Environment. *Dynamics of Continuous, Discrete and Impulsive Systems – Series B: Applications & Algorithms*, 10 (1-3), 81-94.

Kobayashi, S. & Nomizu, K. (1963) *Foundations of Differential Geometry Volume 1*, 2^{nd} Edition, Interscience Publishers.

Lee, J. (2000) *Introduction to Smooth Manifolds*, Springer-Verlag New York.

Malaney, P. (1996) *The Index Number Problem: A Differential Geometric Approach*. PhD Thesis, Harvard University.

Mantegna, R.N. & Stanley, H.E. (2000) *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge, Cambridge University Press.

Marsh, A. (2019) *Gauge Theories and Fiber Bundles: Definitions, Pictures and Results*. 2^{nd} Version (Revised), Working Paper.

Morisawa, Y. (2009) Toward a Geometric Formulation of Triangular Arbitrage – An Introduction to Gauge Theory of Arbitrage. *Progress of Theoretical Physics Supplements*, 178, 209-215.

Munkres, J. (2000) *Topology*, 2^{nd} Edition, Prentice-Hall.

Paolinelli, G. & Arioli, G. (2018) A Path Integral based Model for Stocks and Order Dynamics. *Physica A: Statistical Mechanics and its Applications*, 510, 387-399.

Paolinelli, G. & Arioli, G. (2019) A Model for Stocks Dynamics based on a Non-Gaussian Path Integral. *Physica A: Statistical Mechanics and its Applications*, 517, 499-514.

Piotrowski, E.W. & Sladkowski, J. (2006) Geometry of financial markets—Towards information theory model of markets, *Physica A: Statistical Mechanics and its Applications*, 382 (1), 228-234.

Shreve, S. (2004) *Stochastic Calculus for Finance I: The Binomial Asset Pricing Model*, Springer-Verlag New York.

Smith, A. & Speed, C. (1998) Gauge Transforms In Stochastic Investment Modelling. *AFIR Colloquium*, 8, 445-487.

Smolin, L. (2009) *Time and symmetry in models of economic markets*, Working Paper, Perimeter Institute for Theoretical Physics, Ontario.

Sokolov, A., Kieu, T. & Melatos, A. (2010) A Note on the Theory of Fast Money Flow Dynamics. *European Physical Journal B*, 76 (4), 637-642.

Sornette, D. (1998) Gauge Theory of Finance?. *International Journal of Modern Physics C*, 9 (3), 505-508.

Tranfield, D., Denyer D. & Smart P. (2003) Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management*, 14, 207-222.

Trautman, A. (1981) Fiber Bundles, Gauge Fields, and Gravitation. *General Relativity and Gravitation*, 1, 287-308.

Tu, L. (2011) *An Introduction to Manifolds*, 2^{nd} Edition, Springer-Verlag New York.

Tu, L. (2017) *Differential Geometry: Connections, Curvature, and Characteristic Classes*, Springer International Publishing.

Way, R. (2010) *Introduction to Connections on Principal Fibre Bundles*. Working Paper, University of Surrey.

Young, K. (1999) Foreign Exchange Market as a Lattice Gauge Theory. *American Journal of Physics*, 67 (10), 862-868.

Zhou, S. & Xiao, L. (2010) An Application of Symmetry Approach to Finance: Gauge Symmetry in Finance. *Symmetry – Basel*, 2, 1763-1775.

APPENDIX I. Summary of Chosen Papers

Year	Authors	Journal	Objective of the Paper	Application of Gauge Theory	Conclusions
1998	Sornette	International Journal of Modern Physics C	Criticize early attempts at formulating a gauge theory of finance, proposed by Ilinski in prior work	There is no application of gauge theory per se	Exponential probability weights used *ad hoc*, with no explanation; no reason to assume that complex arbitrage strategies can be built with elementary ones
1999	Young	American Journal of Physics	Conceptualize a simple foreign exchange market as a lattice gauge theory	The idea of gauge symmetry is used, and arbitrage opportunities are built with this invariance principle in mind	Currency units are a matter of convention, and an arbitrary rescaling doesn't change the dynamics underneath. Gains from arbitrage have the geometric structure of a curvature tensor, and as such are invariant under rescaling
2000	Ilinski	Journal of Physics A: Mathematical and Theoretical	Show that a general financial market has the geometric structure of a trivial fibre bundle	Arbitrary changes in asset's units constitute a gauge symmetry of the fibre bundle	General financial markets are fibre bundles; There exists gauge symmetry regarding *numéraire* changes – price rescaling, share splitting or asset's units changing leaves the dynamics unchanged
2002	Kholodnyi	Russian Journal of Mathematical Physics	Introduce the notion of beliefs-preferences gauge symmetry in order to evaluate European contingent claims on a market of a single asset not described by a Markov process	There exists quantities invariant to certain changes in the beliefs and preferences of a market participant (gauge invariant)	For a market of one asset, an equation for the evaluation of European contingent claims is derived, independent of the beliefs or preferences of the trading agents, interpreted as a generalization of Markovian diffusion processes
2002	Kholodnyi	European Physical Journal B			
2003	Kholodnyi	Dynamics of Continuous, Discrete and Impulsive Systems – Series B: Applications & Algorithms			
2009	Morisawa	Progress of Theoretical Physics Supplements	Describe arbitrage as the curvature of the gauge field of a portfolio ratio space	Exchange rates and prices are seen as gauge fields. There exists a gauge symmetry: redenomination invariance	Arbitrage opportunities share similarities with the notion of curved gauge fields

Year	Authors	Journal	Objective of the Paper	Application of Gauge Theory	Conclusions
2010	Dupoyet, Fiebig & Musgrove	Physica A: Statistical Mechanics and its Applications	Model high frequency trading in a general financial market	The probability distribution is acquired by minimizing gains associated with arbitrage and risky strategies, which are gauge-invariant quantities	With a numerical algorithm, it is found that the simulation reproduces the fat tails shape observed in the NASDAQ index historical data, albeit an underestimation of real fat tails
2010	Sokolov, Kieu & Melatos	European Physical Journal B	Examine statements put forth by Ilinski regarding its Gauge Theory of Arbitrage in the book “Physics of Finance”	There is no application of gauge theory per se	Algebraic mistakes were found when deriving the dynamical equations from the Lagrangian, and correcting those leads to unrealistic behaviour of exchange rates. Furthermore, an inconsistency is present in the dynamics derivation
2010	Zhou & Xiao	Symmetry – Basel	Build an option pricing model which is invariant under local *numéraire* transformations	Local *numéraire* transformations are gauge transformations in a fibre bundle, the price-time space as basis and the possible units of the numéraire as fibre	A general option pricing equation can be written as the covariant Laplace equation on the fibre bundle
2012	Dupoyet, Fiebig & Musgrove	Physica A: Statistical Mechanics and its Applications	Introduce an updating mechanism driving the system to a self-organized critical state to the already proposed gauge theory of arbitrage	The model bases itself on gauge theory formalism, similarly to Dupoyet et al. (2010)	Using GARCH(1,1), time series generated with the model present realistic results, agreeing quite well with historical data
2015	Farinelli	Journal of Geometric Mechanics	Integrate stochastic finance in a differential geometric framework (stochastic differential geometry) where arbitrage is allowed	Each financial instrument can be identified with a pair of deflator and term structure (the gauge), and dynamics are invariant under cashflow changes	Stochastic Euler-Lagrange equations are obtained by minimizing arbitrage, describing the markets dynamics in and out of equilibrium for a closed market
2018	Paolinelli & Arioli	Physica A: Statistical Mechanics and its Applications	Numerically solve the probability density function of asset prices proposed by Ilinski for short-term dynamics	The probability density function at hand is constructed with logic from quantum gauge theory	By adding extra perturbations, a better fit with real data is achieved, comparing with prior work on the same topic
2019	Paolinelli & Arioli	Physica A: Statistical Mechanics and its Applications	Model the dynamics of stock prices in such a way that the dynamics are non-Gaussian	The action constructed is achieved by searching for a gauge-invariant functional adjustable to short and long term horizons via three parameters	Constructing a non-quadratic path integral leads to a model well-fitted with real data, for different time horizons, and with only 3 parameters

APPENDIX II. String of Keywords

Keywords†	
Gauge Theory	Finance
(gauge NEAR/0 theor*) OR (gauge NEAR/0 field*) OR (gauge NEAR/0 group*) OR (gauge NEAR/0 transf*) OR (gauge NEAR/0 *symme*) OR (gauge NEAR/0 invar*) OR (gauge NEAR/0 fix*) OR (gauge NEAR/0 indepen*) OR (gauge NEAR/0 boson*) OR	*financ* OR "crisis" OR "great depression" OR market* OR securit* OR "measurement and data" OR portfolio* OR *investment* OR asset* OR capital* OR "diversification" OR "rate of return" OR risk* OR "technical analysis" OR (trading NEAR/0 volume*) OR (interest NEAR/0 rate*) OR annuit* OR "CAPM" OR copula* OR "DAX" OR "Dow" OR equit* OR "January effect" OR "LIBOR" OR "NASDAQ" OR "NYSE" OR (price NEAR/0 earning*) OR "random walk hypothesis" OR stock* OR "T bill" OR "T bond" OR zerobond* OR contingen* OR "futures" OR (option NEAR/0 pric*) OR commodit* OR "Black Scholes" OR cryptocurrenc* OR (forward NEAR/0 pric*) OR "options" OR "put call" OR "event studies" OR insider* OR acquisition* OR announcement* OR "news" OR eurobond* OR "reserves" OR sovereign* OR "Tobin" OR "returns" OR volatil* OR "Basel" OR *regulation* OR investor* OR polic* OR "SEC" OR *bank* OR depositor* OR mortgage* OR "building society" OR credit* OR *debt* OR debit* OR *insurance* OR "fixed term" OR overdraft* OR (reserve NEAR/0 requirement*) OR loan* OR actuar* OR deduct* OR "fully funded" OR "funds" OR "clearing house" OR "fund" OR "REITS" OR broker* OR rating* OR "IPO" OR bailout* OR "Dodd Frank" OR "FDIC" OR federal* OR "FSLIC" OR "Glass Steagall" OR "stress testing" OR corporat* OR firm* OR "animal spirit" OR inventor* OR overcapacit* OR "goodwill" OR amortization* OR "leasing" OR leverag* OR "Modigliani Miller" OR shareholder* OR insolven* OR "liquidation" OR "board behaviour" OR buyout* OR divestiture* OR executive* OR governanc* OR "proxy" OR restructur* OR takeover* OR payout* OR dividend* OR profit* OR "regulatory" OR "Sarbanes Oxley" OR anchor* OR "herd behaviour" OR "hindsight bias" OR "loss aversion" OR "mental accounting" OR overreaction* OR "prospect theory" OR underreaction*

†All keywords were adapted to the database *Web of Science*, by introducing either the operators NEAR/0, * or "". To combine two keywords of the same type to construct a string, the Boolean operator OR was used, and to combine both gauge theoretical and financial strings the operator AND was use.