Silvio Petronzio

A Non-Structural Approach to Option Hedging via Orthogonal Polynomials

Master's degree thesis

LUISS Guido Carli University

July 2020

LUISS GUIDO CARLI UNIVERSITY
Faculty of Economics and Finance
Degree Course in Economics and Finance

A Non-Structural Approach to Option Hedging via Orthogonal Polynomials

Supervisors:
Prof. Paolo SANTUCCI DE MAGISTRIS
Prof. Stefano GRASSI

Silvio Petronzio
Freshman n. 713711

Graduation session: July 2020
Academic Year 2019-2020

Table of Contents

Introduction **5**

1 Orthogonal Polynomials **6**
- 1.1 Elements of Measure theory 7
- 1.2 Construction of the polynomials 11
- 1.3 The orthogonal expansion 15

2 Non-structural hedging **20**
- 2.1 General assumptions 21
- 2.2 Option hedging approach 24
- 2.3 Computation of the Greeks 26
- 2.4 Computation of the risk-neutral moments 28
- 2.5 Choosing the kernel 31

3 Empirical results **34**
- 3.1 The SPX and the VIX : a brief retrospective 34
- 3.2 Analysis of driving risk factors for SPX options 37
- 3.3 Analysis of the risk-neutral moments dependency from the variance swap 44
- 3.4 How to read the RND in turbulent times : from the 2008 crisis to COVID-19 46
- 3.5 Measuring the hedging gain and solving the correlation issues 51
- 3.6 Conclusions 55

A Appendix **56**

Bibliography **58**

Introduction

As the market state becomes ever more susceptible to the aggregate contribute of multiple sources of instability and the proliferation of increasingly sophisticated financial instruments is witnessed, each customizable to suit the diverse adjustment needs, the common practice of minimizing the intrinsic portfolio risk by specular investments, known as hedging, has evolved as a well established and clear-cut praxis, widely included among the retail traders priorities and mandatorily required by regulators to institutional investors. Derivative securities, products whose payoff depends on the trend of an underlying asset, are the principle devices used to channel the loss uncertainty and balance the pool of diversified agent positions. Option contracts, in particular, allow to manage a persistent and effective hedging if properly offset by an opposite dynamic replication of their behaviour, obtained by weighting a collection of stocks entangled with call or put variation. These weight measuring hedge ratios are known as *Greeks* and they match the sensitivity of the option to a determinant pilot risk factor. The most extensively accounted ones are the delta and the vega, corresponding to the partial derivatives of the option price with respect to the level of the underlying and to its volatility. The accuracy of the Greeks estimation is responsible for the hedging quality and it is typically associated with the methodology applied for the extrapolation of the risk-neutral density. This function, whose existence is guaranteed only in arbitrageless venues, is the probability distribution of the prices of the underlying asset in the measure $\mathbb{Q}$, which prices the contingent claims from the perspective of a risk-neutral investor. In the vast and continuously updating literature on the density extraction a first fundamental distinction is made between structural and non-structural models. The former provide a complete description of the stock prices dynamic, the latter derive instead the density by resourcing only from a partial or absent definition of the underlying stochastic process. Non-structural approaches could themselves be classified in parametric, where a direct expression of the risk-neutral density is proposed, and semi or non-parametric, where the density is estimated with approximation techniques. A widespread trading strategy is the so-called practitioners' Black-Scholes

consisting in a delta hedging, with the hedging coefficient computed with the Black-Scholes formula. This is a structural approach as in the Black-Scholes setup the asset price follows a log-normal diffusion process, preserved in the risk-neutral measure $\mathbb{Q}$ for the Girsanov theorem. The main caveats of this procedure stand in the empirically proved failure of the hypotheses, since the volatility of the underlying is not constant and the process may exhibit jumps, and the absence of the vega. More and more refined models have been proposed, with a stochastic volatility and volatility of volatility, but in general any structural or non-structural parametric approach suffers from some recurring drawbacks, especially when historical data are used. When, indeed, a relatively simple model is taken the results could be biased and not adherent to the actual observations, whereas when a complex model is considered, the trade-off is detrimental for the processing time requested for the determination of the parameters.
In this thesis we propose a non-structural semi-parametric approach for the risk-neutral density extraction and the consequent hedging. Our methodology, along the lines of 9, will only rely upon weak regularity conditions without imposing binding constraints for the underlying stochastic process, revisiting and generalizing some standard procedures based on polynomial expansion such as the Edgeworth series or, with the proper adaptation, the Gram-Charlier A series. The Greeks will be then computed from the derived risk-neutral moments via some cunning heuristic identities, later empirically verified with the other assumptions. Beside this main purpose, the work will be set for several additional aims. By venturing in the field of option hedging we will notice how the areas of derivative pricing and economics will intertwine with our findings. We will therefore often detail the whole context by focusing on the VIX puzzle phenomen explanations, outlining how the Greeks should in general change according to the option moneyness and, most importantly, refer to the many ways the risk-neutral density could be utilized in other academic areas. This is done to endow the work with a broader scope, indirectly suggesting some possible variations on the theme treated. The contents of the thesis articulate as follows :

- *Chapter 1*
  We will start by building the mathematical apparatus of our strategy. In particular, our main aim will be to construct the orthogonal polynomials recursively and to study the conditions under which these can allow an infinite series expansion for a function. The exposition is formalized in such a way that any unfamiliar reader can follow its development, starting from elements of Measure Theory and reaching the proof of the completeness of the orthonormal polynomial basis.

- *Chapter 2*
  We will then move to implementing our results in the financial theory, showing how to resort to the polynomials for constructing the hedging. This will bring us to many technical passages since first we will have to estimate the risk-neutral moments, then find a way to express the Greeks as functions of tradable contracts and finally select properly the kernel density.

- *Chapter 3*
  The final chapter will be divided in two halves. The first part will be dedicated to support our previous assumptions against an empirical background, a procedure that will dissect the inner mechanism determining the option variance. The second half will start with a digression on the several applications of the risk-neutral density followed by a guide on how to interpret its form and that of the Greeks and infer the potential future market movements. The final section will describe how to measure the hedging in a time span with periodic rebalancing and how some potential flaws in the strategy, emerging in low volatility times, could be resolved.

1

Orthogonal Polynomials

We will begin by rigorously deriving the theoretical framework underlying the rationale of our hedging strategy. Our focus is to formally establish the conditions under which a probability density of a stock price, itself determining the option payoff, can be expressed as an infinite linear combination of polynomials. We will see that this process parallels a traditional result of Real Analysis which is the construction of a complete basis for an Hilbert space i.e., intuitively, an infinite dimensional vector space. In particular, our aim will be to find an orthonormal basis of polynomials in the vector space of square-integrable functions. We will do this in three steps. First, the unacquainted reader will be introduced to the lexicon of Measure Theory, to the notions that will be extensively employed throughout the rest of the chapter and to the few foundational results our proofs will be subsequently grounded on. We will provide only the necessary elements and thus, for further details, we recommend the reference [2]. Secondly, we will constructively prove the existence of the orthogonal polynomials via a standard orthonormalization procedure which will then pave the way for a recursive determination of their coefficients. Finally, the last section will be devoted to the presentation of the core statement and of the main theorem the proof centers upon.

1.1 Elements of Measure theory

First we need to formally identify the measurable space and the Borel measure. We start with the basic definitions.

Note 1.1. Given a set X we recall the definition of its power set $\mathcal{P}(X) = \{E : E \subseteq X\}$.

Definition 1.1. Given a set X, a familiy of subsets $\mathcal{A} \in \mathcal{P}(X)$ is an *algebra* if :

i) $\varnothing, X \in \mathcal{A}$.

ii) $\forall A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}; A^c \in \mathcal{A}$.

$\mathcal{A}$ is a *σ-algebra* if for any family of subsets $\{A_i\}_{i\in\mathbb{N}} \subseteq X$ we also have

iii) $\bigcup_{i\in\mathbb{N}} A_i \subseteq X$.

The pair $(X, \mathcal{A})$ is then called a *measurable space*.

Definition 1.2. Given two measurable spaces $(X, \mathcal{A}_1)$ and $(Y, \mathcal{A}_2)$,a function $f : X \to Y$ is said to be *measurable* if

$$f^{-1}(E) := \{x \in X : \ f(x) \in E\} \in \mathcal{A}_1, \ \ \forall E \in \mathcal{A}_2 \tag{1.1}$$

Definition 1.3. Given a set X and a subset $\mathcal{F} \in \mathcal{P}(X)$, the *$\sigma$-algebra generated by* $\mathcal{F}$ is the σ-algebra $\mathcal{A}$ such that for any σ-algebra $\mathcal{A}'$ that contains $\mathcal{F}$ we have $\mathcal{A} \subseteq \mathcal{A}'$.

Definition 1.4. Given a non empty set X, a family of subsets $\Sigma \in \mathcal{P}(X)$ is a *topology* on X if

i) $\varnothing, X \in \Sigma$.

ii) $\{A_i\}_{i\in\mathbb{I}} \in \Sigma \Rightarrow \bigcup_{i\in\mathbb{I}} A_i \in \Sigma$.

iii) $A, B \in \Sigma \Rightarrow A \cap B \in \Sigma$.

The pair (X, Σ) is called a *topological space* and the elements of Σ are called *open sets of* X.

Definition 1.5. Given a topological space (X, Σ), the σ-algebra generated by the open sets of X is called *Borel σ-algebra on X.*

Definition 1.6. Given a measurable space $(X, \mathcal{A})$, a *measure* μ on $(X, \mathcal{A})$ is a function $\mu : \mathcal{A} \to [0, +\infty]$ such that

i) $\mu(\varnothing) = 0$.

ii) $\{A_i\}_{i\in\mathbb{N}} \in \mathcal{A} : A_i \cap A_j = \varnothing \, \forall i, j \in \mathbb{N} \Rightarrow \mu(\bigcup_{i=0}^{+\infty} A_i) = \sum_{i=0}^{+\infty} \mu(A_i)$.

The triple $(X, \mathcal{A}, \mu)$ is called *measure space.*

Definition 1.7. Given a measure space $(X, \mathcal{A}, \mu)$, a property P is said to hold *μ-almost everywhere* (abbreviated μ-a.e.) in X if there exists a set $N \in \mathcal{A}$ such that $\mu(N) = 0$ and all $x \in X/N$ satisfy the property P.

Now we need to deliver our first standard result which will implicitly take part to the hypothesis of the core statement : the Radon-Nikodym Theorem.

Definition 1.8. Given two measures μ and ν on a measure space $(X, \mathcal{A})$, μ is said to be *absolutely continuous with respect to* ν if

$$\forall E \in \mathcal{A} \quad \nu(E) = 0 \Rightarrow \mu(E) = 0 \tag{1.2}$$

In which case we will write $\mu << \nu$.

Definition 1.9. A measure μ on a measurable space $(X, \mathcal{A})$ is called *σ-finite* if there exists a family $\{A_n\}_{n\in\mathbb{N}} \in \mathcal{A}$ such that $\mu(A_n) < \infty$ for any $n \in \mathbb{N}$ and $\bigcup_{n\in\mathbb{N}} A_n = X$.

Definition 1.10. A *Borel measure* is any measure μ on a measure space $(X, \mathcal{B}(X))$ with $\mathcal{B}(X)$ Borel σ-algebra on X.

Remark 1.1. Given an interval $(a, b) \subseteq \mathbb{R}$, a Borel measure on $((a, b), \mathcal{B}(a, b))$ with Borel σ-algebra generated by the open intervals of (a, b) is clearly σ-finite. If $a = -\infty$ and/or $b = +\infty$ we take $A_n = (-n, n)$.

Theorem 1.11 (Radon-Nikodym)**.** *Let μ and ν be two σ-finite measures on a measurable space $(X, \mathcal{A})$. If $\mu << \nu$ then there exists a measurable function $f : X \to [0, +\infty)$ such that*

$$\mu(A) = \int_A f \, d\nu \qquad \forall A \subseteq X \tag{1.3}$$

The function f is called Radon-Nikodym derivative and is denoted by $\frac{d\mu}{d\nu}$.

Proof. See [2], Theorem 6.10. □

We can now gradually present the definition of Hilbert space, a pivotal constituent of Real Analysis and a mathematical object that will be essential for our discussion. Additionally, we will provide the definition of closure, whose meaning will prove necessary thereafter to describe a complete basis. Let us start from

Definition 1.12. Given a vector space V over the field F equal to $\mathbb{R}$ or $\mathbb{C}$, a function $\langle .,. \rangle : V \times V \to F$ is an *inner product* if for any $x, y, z \in X$ and $a \in \mathbb{R}$ the following holds:

i) $\langle x, y \rangle = \overline{\langle y, x \rangle}$

ii) $\langle ax, y \rangle = a\langle x, y \rangle, \quad \langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$

iii) $\langle x, x \rangle > 0 \quad x \in V/\{0\}$

The pair $(V, \langle .,. \rangle)$ is then called an *inner product space*.

Definition 1.13. Given a non empty space X, a function $d : X \times X \to \mathbb{R}$ is a *distance on* X if for any $x, y, z \in X$ the following holds :

i) $d(x, y) =\Leftrightarrow x = y$

ii) $d(x, y) = d(y, x)$

iii) $d(x, z) \leq d(x, y) + d(x, z)$

The pair (X, d) is then called a *metric space*.

Remark 1.2. An inner product space is a metric space with the distance given by $d(x, y) = \|x-y\|$, which will be called *induced distance*, and a metric space (X, d) is always a topological space. It is indeed easy to see that the family of sets $B_r(x) = \{y \in X : d(x, y) < r\}$ with $r \in \mathbb{R}$ induces a topology on X.

Definition 1.14. Given a topological space (X, Σ) and a subset $S \subseteq X$ the *closure* of S is the set $\overline{S} = \{x \in X | \forall A \in \Sigma, x \in A \Rightarrow A \cap S \neq \varnothing\}$.

Remark 1.3. Given a metric space (X, d) with the topology induced by the sets $B_r(x)$ and such that for a subset $S \subseteq X$ the equality $\overline{S} = X$ holds, we can always build a sequence of $(x_n)_{n\in\mathbb{N}} \in X$ such that $\lim_{n\to+\infty} d(x_n, x) = 0$. The sequence is built by taking $x_n \in B_{\frac{1}{n}}(x)$ for each $n \in \mathbb{N}$ which exists by the definition of closure.

Definition 1.15. A metric space (X, d) is *complete* if any sequence $(x_n)_{n\in\mathbb{N}} \in X$ such that

$$\forall\, \epsilon > 0\; \exists N\; s.t.\, \forall\; m, n > N\;\, d(x_m, x_n) < \epsilon \tag{1.4}$$

converges in X i.e. $\lim_{n\to+\infty} d(x_n, x) = 0$ for some $x \in X$.

Definition 1.16. An inner product space $(H, \langle .,. \rangle)$ that is also a complete metric space with respect to the distance induced by the inner product is called an *Hilbert space*.

Proposition 1.17 (Bessel's inequality)**.** *Let f_n be a sequence of vectors in an Hilbert space $(H, \langle ., . \rangle)$ such that $\langle f_n, f_m \rangle = \Delta_{nm}$. Then for any f in H we have the Bessel's inequality :*

$$\sum_{k=0}^{+\infty} f_k^2 \leq \|f\|^2 \tag{1.5}$$

Proof. Let $s_n = \sum_{k=0}^{n} \langle f, f_k \rangle f_k$. The following equalities are easy to derive

$$\langle s_n, s_n \rangle = \langle s_n, f \rangle = \sum_{k=0}^{n} \langle f, f_k \rangle^2 \tag{1.6}$$

We then obtain the following

$$\langle s_n - f, s_n - f \rangle = \langle s_n, s_n \rangle - 2\langle s_n, f \rangle + \langle f, f \rangle = \langle f, f \rangle - \sum_{k=0}^{n} \langle f, f_k \rangle^2 \tag{1.7}$$

Since $\langle s_n - f, s_n - f \rangle \geq 0$ we have

$$\sum_{k=0}^{n} \langle f, f_k \rangle^2 \leq \langle f, f \rangle \qquad \forall n \in \mathbb{N} \tag{1.8}$$

□

Theorem 1.18 (Orthogonal decomposition)**.** *Given an Hilbert space $(H, \langle ., . \rangle)$ and a closed vector subspace $M \subseteq H$, we have $H = M \oplus M^{\perp}$ where*

$$M^{\perp} = \{x \in H : \langle x, y \rangle = 0,\ y \in M\}. \tag{1.9}$$

For any $x \in H$ there thus exist two unique vectors $y \in M$ and $z \in M^{\perp}$ such that $x = y + z$.

Proof. See [5], Proposition 4.11. □

The last pillar is the L^p space. Our main concern will be just to prove that L^2 is an Hilbert space but, in order to structure this finding, an overall description of these spaces as well as a complete overview of their nature is strictly required.

Definition 1.19. Given a measure space $(X, \mathcal{A}, \mu)$ and a function $f : X \to \mathbb{R}$, we define the relation $f \rho g \Leftrightarrow f \equiv g$ μ-almost everywhere.

Definition 1.20. Given a measure space $(X, \mathcal{A}, \mu)$ and a real number $p \in [0, \infty]$, we define the sets

$$L^p_\mu(X) = \{f : X \to \mathbb{R} \quad measurable : \|f\|_p = (\int_X |f|^p d\mu)^{\frac{1}{p}} < +\infty\}/\rho \qquad p \in [0, \infty) \tag{1.10}$$

$$L^\infty_\mu(X) = \{f : X \to \mathbb{R} \quad measurable : \|f\|_\infty sup\{b \in \mathbb{R} : \mu(\{x : |f(x)| < b\}) = 0\} < +\infty\}/\rho \tag{1.11}$$

Where ρ denotes the relation of Definition 1.19.
We will call each set an L^p*-space* and the values $\|f\|_p$ and $\|f\|_\infty$ the L^p*-norm* and the L^∞*-norm of* f. If μ is the Lebesgue measure we will just write $L^p(X)$.

Note 1.2. The notation used to define the L^p space indicates that the single elements of $L^p_\mu(X)$ are so called *equivalence classes* i.e. sets $[f]$ for any function f such that $g \in [f]$ if $f \rho g$.

Remark 1.4. Clearly $(L^p_\mu(X), \|.\|_p)$ with $p \in [0, \infty]$ is a metric space.

Theorem 1.21. *Given a measure space* $(X, \mathcal{A}, \mu)$ *such that* $\mu(B) \geq 0\ \forall B \in \mathcal{A}$, *the metric space* $(L^p_\mu(X), \|.\|_p)$ *is complete for every* $p \in [0, \infty]$.

Proof. See [2], Theorem 3.11. □

Remark 1.5. Observe that if $\mu << \nu$ with μ σ-finite and ν satisfying the property of the Theorem, i.e. the positivity, then the Theorem holds also for $(L^p_\mu(X), \|.\|_p)$ since μ is positive for the Radon-Nikodym Theorem. In particular this is true if ν is a Lebesgue measure.

Corollary 1.22. *Given a measure space* $(X, \mathcal{A}, \mu)$ *such that* $\mu(B) \geq 0\ \forall B \in \mathcal{A}$, *the function* $\langle ., . \rangle_\mu : L^2_\mu(X) \times L^2_\mu(X) \to \mathbb{R}$ *such that*

$$\langle f, g \rangle_\mu = \int_X f(x) g(x) d\mu \qquad \forall f, g \in L^2_\mu(X) \tag{1.12}$$

is an inner product and the inner product space $(L^2_\mu(X), \langle ., . \rangle_\mu)$ *is an Hilbert space.*

Proof. It is a consequence of Theorem 1.21 after having easily verified that the properties of Definition 1.12 are satisfied. □

1.2 Construction of the polynomials

To build the polynomials we will use the Gram-Schmidt orthogonalization, a well-known technique of Linear Algebra that consists in extracting an orthogonal basis from a canonical one by subtracting the projection of each vector on the span of the others to the vector itself. We deliver first the

Definition 1.23. Given a Borel measure μ on an interval $(a,b) \subseteq \mathbb{R}$ a sequence of real polynomials $(p_n(x))_{n\in\mathbb{N}}$ with degree $p_n(x) = n$ for each $n \in \mathbb{N}$ is called *orthogonal* on (a,b) with respect to the weight function $\phi : (a,b) \to \mathbb{R}$ if

$$\int_a^b \phi^2(x)p_n(x)p_m(x)d\mu(x) = h_n\Delta_{nm}, \text{ with } \Delta_{nm} = \begin{cases} 0 & \text{if } n \neq m \\ 1 & \text{if } n = m \end{cases} \tag{1.13}$$

If $h_n = 1$ for every $n \in \mathbb{N}$ the sequence of polynomials will be called *orthonormal.*

Proposition 1.24 (Gram-Schmidt orthogonalization)**.** *Let $(H, \langle .,. \rangle)$ be an Hilbert space and let $(v_0, ..., v_n, ...) \in H$ be a collection of linearly indipendent vectors. Then the collection of vectors $(w_0, ..., w_n, ...)$ obtained via the process:*

$$w_0 = v_0, \qquad w_j = v_j - \sum_{h=0}^{j-1} \frac{\langle v_j, w_h\rangle}{\langle w_h, w_h\rangle} w_h \qquad j = 1, 2, ... \tag{1.14}$$

Satisfies the following :

1. *$span(w_0, ..., w_j) = span(v_0, ..., v_j) \qquad \forall j \in \mathbb{N}$*
2. *$\langle w_n, w_m\rangle = 0, \qquad n \neq m \qquad \forall n, m \in \mathbb{N}$*
3. *$\langle w_j, v_j\rangle > 0 \qquad \forall j \in \mathbb{N}$*

Proof. We will proceed by induction on n. For $n = 0$ conditions 1. and 2. are satisfied. If we assume the case $n-1$ then we want to find $w_n = \alpha v_n - \beta_1 w_1 - ... - \beta_{n-1}w_{n-1}$. Condition 1. with $j = n$ requires $\alpha \neq 0$ and condition 2. implies

$$0 = \langle w_n, w_j\rangle = \langle \alpha v_n - \beta_0 w_0 - ... - \beta_{n-1}w_{n-1}, w_j\rangle = \alpha\langle v_n, w_j\rangle - \beta_j\langle w_j, w_j\rangle \tag{1.15}$$

Which results in $\beta_j = \alpha\frac{\langle v_n, w_j\rangle}{\langle w_j, w_j\rangle}$ for $j = 1, 2, ...$ and thus

$$w_n = \alpha[v_n - \sum_{h=0}^{n-1} \frac{\langle v_n, w_h\rangle}{\langle w_h, w_h\rangle} w_h] \tag{1.16}$$

We can take $\alpha = 1$ since $\langle w_n, v_n\rangle > 0$ is verified if and only if $\alpha > 0$. Indeed we have that

$$\langle w_n, v_n\rangle = \alpha[\|v_n\|^2 - \sum_{h=0}^{n-1} \frac{|\langle v_n, w_h\rangle|^2}{\|w_h\|}] = \alpha[\sum_{h=0}^{n} \frac{|\langle v_n, w_h\rangle|^2}{\|w_h\|} - \sum_{h=0}^{n-1} \frac{|\langle v_n, w_h\rangle|^2}{\|w_h\|}] \tag{1.17}$$

Since $v_n \notin span(v_0, ..., v_{n-1}) = span(w_0, ..., w_{n-1})$ we have $\langle w_n, v_n\rangle \neq 0$ and because the term in the parenthesis is positive we must have $\alpha > 0$ to obtain 3. □

Given a weight function $\phi(x)$ with finite moments on (a, b) in the measure μ we will build a basis of orthonormal polynomials from the sequence $(\phi, x\phi, x^2\phi, ...)$ which is clearly a basis for ϕ-weighted polynomials.
Once we know the existence of the Gram-Schimdt procedure, though, we can use a recursive formula to define directly the orthonormal polynomials coefficients.

Theorem 1.25. *A sequence of orthonormal polynomials* $\{p_n(x)\}_{n=0}^{+\infty}$ *satisfies*

$$p_n(x) = A_n[xp_{n-1}(x) + B_n p_{n-1}(x) + C_n p_{n-2}(x)] \qquad n = 2, 3, ... \tag{1.18}$$

where

$$A_n \neq 0, \quad B_n = -\langle xp_n(x), p_n(x)\rangle, \quad C_n = -\langle xp_n(x), p_{n-1}(x)\rangle \quad n = 2, 3, ... \tag{1.19}$$

Proof. The first condition for A_n is obvious. We observe that $p_n(x) - A_n xp_{n-1}(x)$ is a polynomial of degree $\leq n$. Since degree $p_n(x) = n$ for every $n \in \mathbb{N}$ we have

$$p_n(x) - A_n xp_{n-1}(x) = \sum_{k=0}^{n-1} c_k p_k(x) \tag{1.20}$$

From which we derive

$$\langle p_n(x) - A_n xp_{n-1}(x), p_k(x)\rangle = \langle \sum_{k=0}^{n-1} c_k p_k(x), p_k(x)\rangle = c_k \tag{1.21}$$

On the other hand

$$c_k = \langle p_n(x) - A_n xp_{n-1}(x), p_k(x)\rangle = -A_n\langle xp_{n-1}(x), p_k(x)\rangle = -A_n\langle p_{n-1}(x), xp_k(x)\rangle \tag{1.22}$$

Where in the last equality we have used the fact that the product is induced by an integral. For $k < n - 2$ we get degree $xp_k(x) < n - 1$ which implies $\langle p_{n-1}(x), xp_k(x)\rangle = 0$ and thus $c_k = 0$. We therefore have

$$p_n(x) - A_n xp_{n-1}(x) = c_{n-1}p_{n-1}(x) + c_{n-2}p_{n-2}(x) \tag{1.23}$$

The Theorem results from rearranging the equation and gathering A_n and the formulas for B_n and C_n can be obtained by taking the respective scalar products. □

We can now fully derive the coefficients for the orthonormal polynomials. We start by defining the s-th moment of ϕ^2 by μ which are required to be finite i.e.

$$\mu_s = \int_a^b x^s \phi^2(x) d\mu(x) < \infty \tag{1.24}$$

We then define $w_{i,j}$ as the j-th coefficient of the i-th polynomial. To build the sequence of $(p_n(x))_{n\in\mathbb{N}}$ we orthonormalize $(1, x, x^2, ...)$ in the measure $\phi^2\mu$. The first two polynomials are derived by applying Gram-Schmidt directly :

$$p_0(x) = w_{0,0} = 1 \tag{1.25}$$

$$p_1(x) = w_{1,0} + w_{1,1}x = \frac{x - \frac{\mu_1}{\mu_0}}{(\mu_2 - \frac{\mu_1^2}{\mu_0})^{1/2}} \tag{1.26}$$

For $n \geq 2$ we will use the recursive formula 1.18 :

$$p_n(x) = w_{n,0} + ... + w_{n,n}x^n = A_n[(x + B_n) \sum_{k=0}^{n-1} w_{n-1,k}x^k + C_n \sum_{k=0}^{n-2} w_{n-1,k}x^k] \tag{1.27}$$

Where we can write the recursive coefficients as

$$B_n = -\sum_{k=0}^{n-1}\sum_{q=0}^{n-1} w_{n-1,k}w_{n-1,q}\mu_{k+q+1}, \qquad C_n = -\sum_{k=0}^{n-1}\sum_{q=0}^{n-2} w_{n-1,k}w_{n-2,q}\mu_{k+q+1} \tag{1.28}$$

Since A_n is just a normalization costant we start to define the coefficient $w'_{i,j} = A_n w_{i,j}$ of orthogonal but not orthonormal polynomials which can be obtained by matching the terms in the previous equation, this means that as $i = n$ varies we have

$$w'_{i,j} = \begin{cases} B_n w_{n-1,0} + C_n w_{n-2,0} & \text{if } j = 0, \\ w_{n-1,j-1} + B_n w_{n-1,j} + C_n w_{n-2,j} & \text{if } j = 1, ..., n-2, \\ w_{n-1,n-2} + B_n w_{n-1,n-1} & \text{if } j = n-1, \\ w_{n-1,n-1} & \text{if } j = n, \\ 0 & \text{if } j > n. \end{cases} \tag{1.29}$$

Then A_n can be computed as the normalization term i.e.

$$A_n = \pm(\sum_{k=0}^{n}\sum_{q=0}^{n} w'_{n,k}w'_{n,q}\mu_{k+q})^{-1/2} \tag{1.30}$$

Now that the recursive form for the construction of the orthonormal polynomials coefficients is derived we need to show that this is an actual basis for a probability density i.e. that this can be expressed as an infinite sum of the polynomials.

1.3 The orthogonal expansion

We are finally in the position to state our core Proposition which specifies the assumptions for the existence of the orthogonal expansion and, consequently, of our hedging approach.

Proposition 1.26. *Given a Borel measure μ on an open set $S \subseteq \mathbb{R}$ let the kernel $\phi : S \to \mathbb{R}$ and the target $f : S \to \mathbb{R}$ be two measurable functions on $(S, \mathcal{B}(S), \mu)$ with supp$(f) \subseteq$ supp$(\phi) \subseteq S$ and such that :*

i) The kernel ϕ is different from zero almost everywhere and it satisfies

$$\int |x|^k \phi^2(x) d\mu(x) < \infty, \qquad \forall\, k \in \mathbb{N} \tag{1.31}$$

ii) The target f belongs to the space L^2_μ, i.e.

$$\int f^2(x) d\mu(x) < \infty \tag{1.32}$$

Then the following holds :

1. *There exists a family of polynomials $(p_k)_{k\in\mathbb{N}}$ such that the corresponding ϕ-weighted family $(\phi p_k)_{k\in\mathbb{N}}$ is an orthonormal set in L^2_μ, i.e.*

$$\langle \phi p_k, \phi p_l \rangle_{2,\mu} = \Delta_{kl} \qquad \forall\, k, l \in \mathbb{N} \tag{1.33}$$

2. *The Fourier coefficients are well defined, i.e.*

$$c_k = \langle f, \phi p_k \rangle_{2,\mu} < \infty \qquad \forall\, k \in \mathbb{N} \tag{1.34}$$

3. *The sequence of the pseudo-densities*

$$f_n(x) = \phi(x) \sum_{k=0}^{n} c_k p_k(x) \qquad n \in \mathbb{N} \tag{1.35}$$

   converges in the space L^2_μ.

If μ is absolutely continuous with respect to the Lebesgue measure and if, given the Radon-Nikodym derivative $\frac{d\mu}{dx}$, the following condition holds :

$$\exists\, \alpha > 0 \quad s.t. \quad \frac{d\mu}{dx}\phi^2(x) = \mathcal{O}(e^{-\alpha|x|}) \quad as \quad |x| \to +\infty \tag{1.36}$$

4. *The pseudo-densities f_n converge to the target f in norm, i.e.*

$$\lim_{n\to+\infty} \|f - f_n\|_{2,\mu} = 0 \tag{1.37}$$

Note 1.3. We recall that $f(x) = \mathcal{O}(g(x))$ as $|x| \to +\infty$ if and only if there exist two positive real numbers M and x_0 s.t. $|f(x)| \le Mg(x)$ when $|x| \ge x_0$.

Proof.

1. Use the procedure outlined in Section 1.2 to build an orthonormal set of ϕ-weighted polynomials $(\phi p_k)_{k\in\mathbb{N}}$ from the ϕ-weighted basis of polynomials $\phi, x\phi, x^2\phi, ...$ which would then be well defined for condition i).

2. Use the Bessel's inequality of Proposition 1.17 and condition ii) to get

$$c_k \le (\sum_{k=0}^{+\infty} |c_k|^2)^{1/2} \le \|f\|_{2,\mu} < +\infty \tag{1.38}$$

3. Use the Bessel's inequality of Proposition 1.17 and condition ii) to get

$$\lim_{n\to+\infty} \|f_n\|_{2,\mu} = (\sum_{k=0}^{+\infty} |c_k|^2)^{1/2} \le \|f\|_{2,\mu} < +\infty \tag{1.39}$$

4. If $\mu << \lambda$, with λ Lebesgue measure, by Theorem 1.11 we have $d\mu = \psi dx$ with $\psi : S \to \mathbb{R}^+$ measurable function. Then the family $h_k = \sqrt{\psi}\phi p_k$, for every $k \in \mathbb{N}$, is an orthonormal set in $L^2(S)$. Since we also have that $\|g\|_{2,\mu} = \|\sqrt{\psi} g\|_2$, it is sufficient to prove that for any $g \in L^2(S)$ the sequence

$$g_n(x) = \sum_{k=0}^{n} \langle g, h_k\rangle h_k(x) \qquad n \in \mathbb{N} \tag{1.40}$$

converges in norm to g. Then by defining $g = \sqrt{\psi} f$ we would have

$$\lim_{n\to+\infty} \|f - f_n\|_{2,\mu} = \lim_{n\to+\infty} \|g - g_n\|_2 = 0 \tag{1.41}$$

This fact is proven by the following two Theorems.

□

Definition 1.27. Given an Hilbert space H we will say that an orthonormal set $V \subseteq H$ is *complete* if $\overline{span(V)} = H$ except for almost everywhere null functions.

To prove the following result we will make use of standard Real and Complex Analysis theorems, all of them listed with the respective reference in the final Appendix.

Theorem 1.28 (Hewitt (1954)). *Let* $-\infty \leq a < b \leq +\infty$.*Let* $p(x) \in L^2(a,b)$ *be different from zero a.e. and such that* $p(x) = \mathcal{O}(e^{-\alpha|x|})$ *for some* $\alpha > 0$ *as* $|x| \to +\infty$. *Then the family* $(h_k)_{k\in\mathbb{N}}$ *of orthonormal polynomials formed by applying the Gram-Schmidt process on the set* $\{x^n p(x)\}_{n\in\mathbb{N}}$ *is a complete set in* $L^2(a,b)$.

Note 1.4. We recall, given a vector space $V \subseteq L^2(a,b)$, the definition of its orthogonal complement $V^\perp := \{f \in L^2(a,b) : \langle f, v\rangle_2 = 0 \quad \forall v \in V\}$.

Note 1.5. Given a vector space $V \subseteq L^2(a,b)$, we will write $V = \{0\}$ to mean that any f in V is null almost everywhere.

Proof. Let $B = \{h_1, h_2, ..., h_n, ...\}$ be the collection of orthonormal polynomials. By Theorem 1.18 we have $L^2(a,b) = \overline{span(B)} \bigoplus (\overline{span(B)})^\perp$. Our aim is to show that $\overline{span(B)}^\perp = \{0\}$ as this would imply the completeness of $(h_k)_{k\in\mathbb{N}}$. Let f be an element of $(\overline{span(B)})^\perp$. We have

$$\int_a^b x^n p(x) f(x) dx = 0 \qquad \forall n \in \mathbb{N} \tag{1.42}$$

Let us then define the complex function

$$F(z) = \int_a^b e^{izx} p(x) f(x) dx \tag{1.43}$$

We now want to find the domain of differentiability and thus, by Theorem A.1, analiticity of $F(z)$. Let us first write

$$F(u+iv) = \int_a^b e^{-vx} cos(ux) p(x) f(x) dx + i \int_a^b e^{-vx} sin(ux) p(x) f(x) dx = g(z) + ih(z) \tag{1.44}$$

If a and b are finite we can use the Leibniz rule A.6 to differentiate under the integral sign and show that the Cauchy-Riemann equations of Theorem A.5 are satisfied.This implies that $F(z)$ is analytic in $\mathbb{C}$. If $a = -\infty$ and/or $b = +\infty$ we show that, calling $w(u,v,x)$ the integrand of $g(z)$, for any $|v| < \alpha$ we have

$$|w_v(x)| = |-xe^{-vx} cos(ux) p(x) f(x)| \leq |xe^{-vx} e^{-\alpha|x|}||e^{\alpha|x|} p(x)||f(x)| \tag{1.45}$$

The first term is in $L^1(a,b)$ since $|v| < \alpha$. The second term is bounded by hypothesis for $|x| \geq x_0$ with x_0 s.t. $|e^{\alpha|x|} p(x)| \leq M e^{\alpha|x|} e^{-\alpha|x|} = M$ for some

$M \in \mathbb{R}$. The third term is bounded almost everywhere for $|x|$ greater than some x_1 since $f(x) \in L^2(a,b)$. Setting $x_2 = max(x_0, x_1)$ we have

$$g(z) = \int_{-\infty}^{+\infty} w(u,v,x)dx = \int_{-\infty}^{-x_2} w(u,v,x)dx + \int_{-x_2}^{x_2} w(u,v,x)dx + \int_{x_2}^{+\infty} w(u,v,x)dx \tag{1.46}$$

We always differentiate under the integral for the central one whereas for the remaining we use Theorem A.2 with the conditions we have showed. Applying the same reasoning to $w_u(x)$ and $h(z)$ we see that the Cauchy-Riemann equations are true for $|v| < \alpha$ which is thus where $F(z)$ is analytic. In this domain the n-th derivatives are

$$F^{(n)}(z) = i^n \int_a^b e^{izx} x^n p(x) f(x) dx \qquad \forall n \in \mathbb{N} \tag{1.47}$$

From 1.42 we derive that

$$F^{(n)}(0) = i^n \int_a^b x^n p(x) f(x) dx = 0 \qquad \forall n \in \mathbb{N} \tag{1.48}$$

By Theorem A.3 we have that $F(z) = 0$ identically in its domain of differentiability. In particular we have $F(u) = \int_a^b e^{iux} x^n p(x) f(x) dx = 0$ for any real u. $F(u)$ corresponds to the Fourier transform of $p(x)f(x)$ which is in $L^1(a,b)$ for the inequality of Proposition A.7

$$|\langle |p(x)|, |f(x)| \rangle| \leq \|p(x)\| \|f(x)\| < +\infty \tag{1.49}$$

Where the inner product denotes the integral. We therefore have a function in $L^1(a,b)$ whose Fourier transform vanishes identically and thus by Theorem A.4 we derive that $p(x)f(x) = 0$ almost everywhere. Given that $p(x) \neq 0$ almost everywhere by hypothesis, this implies that $f(x) = 0$ almost everywhere i.e. $(\overline{span(B)})^{\perp} = \{0\}$. □

Remark 1.6. If (a,b) is a finite interval the condition of exponential decay is actually not necessary. This means that if $S \subseteq \mathbb{R}$ is a finite connected set condition 1.36 is not needed to prove the completeness.

Setting $p(x) = \sqrt{\psi(x)}\phi(x)$ we can apply the previous Theorem restricted to the support of ϕ. A linear combination of polynomials will thus converge to any f s.t. $supp(f) \subseteq supp(\phi)$. To complete the proof of Proposition 1.26 we just need the following

Theorem 1.29. *Suppose that for a function f and an orthonormal sequence $\{f_k\}_{k=1}^{+\infty}$ both in an Hilbert space $(H, \langle .,. \rangle)$ we have* $\lim_{n\to+\infty} \|f - \sum_{k=0}^{n} c_k f_k\| = 0$. *Then $c_k = \langle f, f_k \rangle$ for each k.*

Proof. Let $s_n = \sum_{k=0}^{n} c_k f_k$. Fix an m and by Proposition A.7 we have

$$|\langle s_n, f_m\rangle - \langle f, f_m\rangle| = |\langle s_n - f, f_m\rangle| \leq \|s_n - f\|\|f_m\| = \|s_n - f\| \tag{1.50}$$

Taking the limits we get

$$c_m = \lim_{n\to+\infty} \langle s_n, f_m\rangle = \langle f, f_m\rangle. \tag{1.51}$$

□

2
Non-structural hedging

After having articulated in full the mathematical architecture behind our main result, we can now move to the central part where the intuition of the methodology, its formal implementation and the heuristics applied to estimate the pivotal hedging entities will be explained and carefully derived. This chapter will be dedicated to expose, dissect and clarify the strategy in all its technical aspects and it is complementary to the following where the theoretical statements will be matched and supported by their empirical counterpart. Our main idea is to reintroduce a traditional non-structural pricing practice in new terms, to exactly calibrate European options for a hedging portfolio. To provide the necessary context and remark the weight of this method in the existing literature, we will first introduce the reader to the general assumptions as well as to the basic theorems of derivative pricing. Then, we will outline our approach, display its advantages within the structural and non-structural distinction and see how it resources from the orthogonal expansion of the previous chapter. To preserve the model independence we will compute the Greeks and the hedge ratios via some smart approximations expressed and proved in the third section. The last step will consist in finding an efficient way to extract the moments of the underlying density which, to overcome some multicollinearity issues, will be done by resorting to a Principal Component Regression. A final discussion will be dedicated to the identification of the kernel function, another essential constituent of the strategy, whose proper choice will prove determinant for the accuracy of all the relations presented.

2.1 General assumptions

Our strategy will work in a market model satisfying a set of specific hypotheses which will be made explicit and formalized in this section. Beside setting the firm background for the development of our method, this passage will also help to briefly familiarize with the foundations of option pricing. Let us then provide first some well-known and less well-known notions in a rigorous form.

Definition 2.1. A *stochastic process* on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $\omega \to X(t, \omega)$ such that for any fixed t the mapping $\omega \to X(t, \omega)$ is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 2.2. An *arbitrage* is a portfolio whose value at time t is described by a stochastic process X_t such that for some $t + \tau > 0$ the following holds:

$$X_0 = 0, \quad \mathbb{P}\{X_{t+\tau} \geq 0\} = 1, \quad \mathbb{P}\{X_{t+\tau} > 0\} > 0 \tag{2.1}$$

A market with no arbitrage will be called *arbitrage-free*.

Note 2.1. We will often implicitly identify the securities and the portfolios with their stochastic processes.

We will only adopt dynamic trading strategies that do not imply exogenous injections or withdrawals of money. This means that to purchase or sell additional amounts of the existing stocks we need to rebalance our portfolio. More specifically

Definition 2.3. Given a vector of stocks $S_t = (S_t^1, ..., S_t^n)$ and a bond B_t with respective weights $x_t = (x_t^1, ..., x_t^n)$ and y_t, the portfolio $P_t = x_t S_t + y_t B_t$ is said to be *self-financing* if

$$S_t dx_t + dS_t dx_t + B_t dy_t + dB_t dy_t = 0 \tag{2.2}$$

The definition could be better understood if we assume that S_t and B_t change values in discrete times i.e. if for any $t \in \mathbb{N}$ and $t^+ \in [t, t+1)$ we have $S_t = S_{t^+}$ and $B_t = B_{t^+}$. Condition 2.2 would then be rewritten only for rebalancing times as

$$S_t(x_{t+1} - x_t) + B_t(y_{t+1} - y_t) = 0 \Rightarrow P_t = P_{t+1} \tag{2.3}$$

This means that, as the portfolio components remain constant in the time interval, its value before changing the weights should equal its value afterwards and thus there are no external money infusions.

Definition 2.4. A market is *complete* if for every derivative security V_t it is possible to create a portfolio H_t made only of riskless assets and stocks such that at the derivative maturity time $t + \tau$ we have $V_{t+\tau} = H_{t+\tau}$.

Definition 2.5. A *filtered probability space* $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathcal{F}, \mathbb{P})$ is a probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ with a *filtration* i.e. a family of σ-algebras $(\mathcal{F}_t)_{t\geq 0}$ such that $\mathcal{F}_t \subseteq \mathcal{F}$ and $\mathcal{F}_s < \mathcal{F}_t$ for $s < t$. If the following holds :

i) $\forall N \in \mathcal{F}\ s.t.\ \mathbb{P}(N) = 0 \Rightarrow N \in \mathcal{F}_0$.

ii) $\mathcal{F}_t \subset \mathcal{F}_s$ for $t \leq s$ and $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$.

we will say that $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathcal{F}, \mathbb{P})$ *satisfies the usual conditions.*

Definition 2.6. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra and X be ab either nonnegative or integrable random variable. The *conditional expectation of X given $\mathcal{G}$*, denoted $\mathbb{E}[X|G]$, is any random variable that satisfies

i) $\mathbb{E}[X|G]$ is measurable on $(\Omega, \mathcal{G})$.

ii) $\int_A \mathbb{E}[X|G](\omega) d\mathbb{P}(\omega) = \int_A X(\omega) d\mathbb{P}(\omega), \quad \forall A \in \mathcal{G}$.

Definition 2.7. Given a filtered probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathcal{F}, \mathbb{P})$, a *martingale* with respect to $\mathbb{P}$ is a real-valued stochastic process M_t on $(\Omega, \mathcal{F}, \mathbb{P})$ such that

i) M_t is $\mathcal{F}_t$*-adapted* i.e. it is a measurable function on $(\Omega, \mathcal{F}_t)$ for every $t \geq 0$.

ii) $M_t \in L^1(\mathbb{P}) \quad \forall t \geq 0$.

iii) $\mathbb{E}[M_t|\mathcal{F}_s] = M_s$ $\mathbb{P}$-almost everywhere.

Definition 2.8. Two σ-finite measures $\mathbb{P}$ and $\mathbb{Q}$ are *equivalent* if $\mathbb{P} << \mathbb{Q}$ and $\mathbb{Q} << \mathbb{P}$.

We will work with European derivative securities i.e. contingent claims that can be exercised only at maturity. This is because the prominent literature of structural and non-structural pricing methods, starting from the Black-Scholes equation, is mainly centered on this kind of options. We will always denote with Z_t the value of a generic underlying asset at a time $t \geq 0$ which will be a stochastic process on the filtered probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathcal{F}, \mathbb{P})$. we will set $t + \tau$ as the time of expiration for the contingent claim on this underlying. We can now express the general assumptions we will be working with :

1. The market is arbitrage-free.
2. The market is complete.
3. The market is frictionless.
4. The risk-free interest rates are set to zero without loss of generality.
5. The price of any derivative security on the underlying Z_t depends on a finite number of traded risky factors $\xi_{t,\tau} = (\xi^1_{t,\tau}, ..., \xi^q_{t,\tau})$.
6. $\exists\, n \geq 2\ s.t.\ \mathbb{E}^{\mathbb{P}}[Z^k_{t+\tau}] < \infty,\ k = 0, 1, ..., n.$

Assumptions 1. and 2. are at the basis of option pricing and they are necessary to prove the existence and uniqueness of an equivalent martingale measure $\mathbb{Q}$ via, respectively, the following two theorems :

Theorem 2.9 (First fundamental theorem of asset pricing)**.** *The market is arbitrage-free if and only if there exists a probability measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ such that the discounted price process of every tradable asset is a martingale with respect to $\mathbb{Q}$.*

Proof. See [6], Theorem 5.4.7. □

We will only work with the measure $\mathbb{Q}$ which is called *risk-neutral probability measure.*

Theorem 2.10 (Second fundamental theorem of asset pricing)**.** *The market is complete if and only if there exists a unique risk-neutral probability measure.*

Proof. See [6], Theorem 5.4.9. □

Our assumptions guarantee thus the existence of a unique risk-neutral probability measure such that for any derivative security on the underlying Z_t with maturity $t + \tau$, payoff function $\Psi(Z_{t+\tau})$ and value $V^{\Psi}_{t,\tau}$ we have

$$V^{\Psi}_{t,\tau}(\xi_{t,\tau}) = E^{\mathbb{Q}}[e^{-r\tau}\Psi(Z_{t+\tau})|\mathcal{F}_t] = E^{\mathbb{Q}}[\Psi(Z_{t+\tau})|\mathcal{F}_t] \tag{2.4}$$

Where $\xi_{t,\tau}$ is the vector of risky factors and in the last equality we have used the hypothesis that $r = 0$. We will denote with $f^{\mathbb{Q}}_{t,\tau}$ the *risk-neutral density*, or RND, i.e. the density of $Z_{t+\tau}$ in $\mathbb{Q}$. Due to Theorem 1.11 and assumption 6., by taking the Radon-Nikodym derivative $\frac{d\mathbb{Q}}{d\mathbb{P}}$ we have that $\mathbb{E}^{\mathbb{Q}}[Z^k_{t+\tau}] < \infty,\ k = 0, 1, ..., n$ for $\mathbb{Q}$ as well.

2.2 Option hedging approach

We can now explicitly characterize our hedging strategy. According to a well established distinction in the literature of financial mathematics, the approach we will develop falls in the category of non-structural methods. The distinction between structural and non-structural approaches to estimate the RND lies at the heart of option pricing and it is extensively treated in Chapter 11 of 7. A structural model delivers a full description of the underlying stochastic dynamics, usually specifying the process of its stochastic volatility as well. A non-structural model provides instead an estimation of the RND without completely describing the process of the underlying value. The fertile field of non-structural methods originates from the Breeden and Litzenberger formula, subsequently benefiting from a consistent stream of contributions. Its advantages when applied to hedging usually rely on the very few assumptions required for the form of the density and the fact that the hedge ratios are not biased, which is instead one troubling issues concerning structural models.
Before explaining our strategy let us clarify what we mean for hedging. Given an option $V^{\Psi}_{t,\tau}$, our aim is to minimize its embedded risk with the following hedging portfolio:

$$P_{t,\tau} = V^{\Psi}_{t,\tau} + \pi^1_{t,\tau}\xi^1_{t,\tau} + ... + \pi^q_{t,\tau}\xi^q_{t,\tau} + \pi^B_{t,\tau}B_{t,\tau}, \qquad \pi^i_{t,\tau} = -\frac{\partial V^{\Psi}_{t,\tau}}{\partial \xi^i_{t,\tau}},\ i = 1, ..., q \tag{2.5}$$

Where $\pi^B_{t,\tau}$ is a position on a risk-free asset $B_{t,\tau}$ chosen to keep the portfolio self-financing. The hedge ratios $\pi^i_{t,\tau}$ correspond to the Greeks of the option and they weight the contracts to make the portfolio constant with respect to the risky factors. We indeed have $\frac{\partial P_{t,\tau}}{\partial \xi^i_{t,\tau}} = 0$ for every $i = 1, ..., q$, which makes the portfolio riskless. We will compute these values by cleverly resorting to the orthogonal polynomials theory derived in Chapter 1, an application that will display a striking non-structural relation between the option and the moments Greeks. Let us start by expressing the RND via the orthogonal expansion of equation 1.35. Since $f^{\mathbb{Q}}_{t,\tau}$ clearly satisfies the hypothesis of Proposition 1.3, by choosing a proper kernel ϕ, later specified in Section 2.5, and the measure $\mu = \phi^{-1}$ we can write

$$f^{\mathbb{Q}}_{t,\tau}(x) \approx f_n(x) = \phi(x)\sum_{k=0}^{n}\langle f^{\mathbb{Q}}_{t,\tau}, \phi p_k\rangle_{2,\mu} p_k(x) = \phi(x)\sum_{k=0}^{n}[\int_0^{+\infty} f^{\mathbb{Q}}_{t,\tau}(x)p_k(x)dx]p_k = \tag{2.6}$$

$$= \phi(x)\sum_{k=0}^{n}\sum_{i=0}^{k}\sum_{j=0}^{k} w_{i,k}M^i_{t,\tau}w_{j,k}x^j = \phi(x)\sum_{k=0}^{n}M^k_{t,\tau}\left(\sum_{i=k}^{n}\sum_{j=i}^{n} w_{i,k}w_{j,k}x^j\right) \tag{2.7}$$

Where $M_{t,\tau}^k$ is the k-th risk-neutral moment $\mathbb{E}^{\mathbb{Q}}[Z_{t+\tau}^k|\mathcal{F}_t]$ and the last equality can be easily proved by induction. Using the martingale property of $\mathbb{Q}$, we also notice that

$$V_{t,\tau}^{\Psi} = \mathbb{E}^{\mathbb{Q}}[\Psi(x)|\mathcal{F}_t] = \int_{\mathbb{R}+} \Psi(x) f_{t,\tau}^{\mathbb{Q}}(x)dx \tag{2.8}$$

Replacing $f_{t,\tau}^{\mathbb{Q}}$ with its truncated expansion 2.7 we have

$$V_{t,\tau}^{\Psi} \approx H_0^{\Psi} + M_{t,\tau}^1 H_1^{\Psi} + ... + M_{t,\tau}^n H_n^{\Psi}, \quad H_k^{\Psi} = \sum_{i=k}^{n} \sum_{j=i}^{n} w_{i,k} w_{j,k} \int_0^{+\infty} x^j \Psi(x)\phi(x)dx \tag{2.9}$$

The importance of the previous equation relies on the fact that the coefficients H_k^{Ψ} do not depend neither on t nor on the risk-neutral moments. Introducing the vectors $M_{t,\tau} = [1, M_{t,\tau}^1, ..., M_{t,\tau}^n]'$ and $H^{\Psi} = [H_0^{\Psi}, ..., H_n^{\Psi}]$ we can express the equation in a compact form as $V_{t,\tau}^{\Psi} = HM_{t,\tau}$. Since ϕ does not change with t either, we can observe that

$$\frac{\partial V_{t,\tau}^{\Psi}}{\partial t} \approx H^{\Psi} \frac{\partial M_{t,\tau}}{\partial t} \Rightarrow \frac{\partial V_{t,\tau}^{\Psi}(\xi_{t,\tau})}{\partial \xi_{t,\tau}} \approx H^{\Psi} \frac{\partial M_{t,\tau}(\xi_{t,\tau})}{\partial \xi_{t,\tau}} \tag{2.10}$$

Where $\frac{\partial V_{t,\tau}^{\Psi}}{\partial \xi_{t,\tau}}$ is a gradient and $\frac{\partial M_{t,\tau}(\xi_{t,\tau})}{\partial \xi_{t,\tau}}$ a Jacobian matrix. If $V_{t,\tau}^{\Psi}$ is a call option $C_{t,\tau}^K$ with strike K, we have $\Psi(x) = (x - K)^+$ and we can denote with H^K the coefficients H^{Ψ} since the payoff only depends on K. For a given set of strikes $K_1, ..., K_p$ we can then consider the vector of respective call options $C_{t,\tau} = [C_{t,\tau}^{K_1}, ..., C_{t,\tau}^{K_p}]'$ and the $p \times n + 1$ matrix H with components $H_{i,j} = H_j^{K_i}$, to write the following approximations

$$C_{t,\tau} \approx HM_{t,\tau}, \qquad \frac{\partial C_{t,\tau}}{\partial \xi_{t,\tau}} \approx H \frac{\partial M_{t,\tau}}{\partial \xi_{t,\tau}} \tag{2.11}$$

The second equality, as well as equation 2.9, delivers a striking relation between the Greeks of the option, i.e. the negative hedge ratios, and the Greeks of the risk-neutral moments. We observe indeed that, since the matrix of coefficients H does not depend on the vector $\xi_{t,\tau}$, the sensitivity of the option from the risk factors is all determined by $M_{t,\tau}$ whereas, for the same reason, the relation with the payoff Ψ is solely ascribable to H. The dependance of $C_{t,\tau}$ from the payoff function and the risk factors is thus decoupled in two terms. This turns out to be the significant part of our strategy because the creation of the hedging portfolio now all boils down to find a functional form to express the moments in terms of tradable contracts. If the moments of the underlying were for instance directly traded as derivative contracts we could immediately use the right member of relation 2.9 to complete our portfolio and take the coefficients H_i^{Ψ} as the hedging ratios. Nevertheless this is not the case, as the risk-neutral moments are not tradable. We will now see that a more subtle relation could be found, connecting the risk-neutral moments to the futures and the variance swaps of the underlying, which will be then proven to be the main driving risk factors.

2.3 Computation of the Greeks

The two observable risk factors that we propose as main determinants for the behaviour of the risk-neutral density are the futures on the underlying and the variance swap. Here follows an explicit definition

Definition 2.11. Given an asset Z_t, the derivative contracts with maturity $t+\tau$ and values

$$F_{t,\tau} = \mathbb{E}^{\mathbb{Q}}[\Psi_1(Z_{t+\tau})] = \mathbb{E}^{\mathbb{Q}}[Z_{t+\tau}|\mathcal{F}_t]] \tag{2.12}$$

$$VS_{t,\tau} = \mathbb{E}^{\mathbb{Q}}[\Psi_2(Z_{t+\tau})] = -\frac{2}{\tau}\mathbb{E}^{\mathbb{Q}}[\log\left(\frac{Z_{t+\tau}}{F_{t,\tau}}\right)|\mathcal{F}_t] \tag{2.13}$$

are respectively called its *future* and its *variance swap*. Their Greeks are defined as

$$\Delta_{t,\tau}^K := \frac{\partial C_{t,\tau}^K}{\partial F_{t,\tau}} \qquad \nu_{t,\tau}^K := \frac{\partial C_{t,\tau}^K}{\partial VS_{t,\tau}} \tag{2.14}$$

and they are respectively called *delta* and *variance swap vega*.

These risk factors are tradable as respectively exchange-traded and over-the-counter derivatives and the intuition behind their choice as candidates is that they best incorporate the fundamental parameters of a density, as well as absorbing the influence of the other non observable moments. The futures is the first and only directly tradable moment of the underlying and it determines the position of the risk-neutral density. Its related Greek, the delta, is the primary almost compulsory ratio adopted by any practitioner of a hedging strategy.The variance swap corresponds approximately to the logarithm of the variance multiplied by a notional and its contribution to the portfolio is manifold. On the one hand it allows to include the volatility parameter in the hedging strategy. On the other, the fact that its payoff is a smooth function results in the contract being sensitive to the change in the following k-th risk-neutral moments. We indeed observe, via a Taylor development of the k-th order, that

$$\log\left(\frac{Z_{t+\tau}}{F_{t,\tau}}\right) \approx \sum_{i=0}^{k} a_i^k \left(\frac{Z_{t+\tau}}{F_{t,\tau}}\right)^i \tag{2.15}$$

This means that the variance swap vega will carry information on the subsequent risk-neutral moments therefore including potential unobservable factors, something that another instrument on the volatility, e.g. a VIX derivative, would have not guaranteed us. As mentioned in the previous section, to compute the Greeks of the option we need to derive a functional relation connecting the contracts to the moments. In the following Proposition we introduce an efficient approximation that does not rely on any modeling assumption.

Proposition 2.12. *Given an asset Z_t, let $M^k_{t,\tau}$ be its k-th risk-neutral moment, $F_{t,\tau}$ its future and $VS_{t,\tau}$ its variance swap. Given the coefficients*

$$\beta_k = \frac{1}{VS_{t,\tau}} \log\left(\frac{M^k_{t,\tau}}{F^k_{t,\tau}}\right), \quad \forall k \in \mathbb{N} \tag{2.16}$$

We have that the following approximation holds :

$$M^k_{t,\tau} \approx F^k_{t,\tau} e^{\beta_k VS_{t,\tau}}, \quad \forall k \in \mathbb{N} \tag{2.17}$$

Proof. For a given k, let us first express the logarithm in the variance swap definition with a Taylor expansion

$$\log\left(\frac{Z_{t+\tau}}{F_{t,\tau}}\right) \approx \sum_{i=0}^{k} a^k_i \left(\frac{Z_{t+\tau}}{F_{t,\tau}}\right)^i \tag{2.18}$$

Where the coefficients a^k_i satisfy the relation

$$\sum_{i=0}^{k} a^k_i \approx \log(1) = 0 \tag{2.19}$$

We will then prove by strong induction that

$$M^k_{t,\tau} \approx F^k_{t,\tau}(1 + \beta_k VS_{t,\tau}), \quad \forall k \in \mathbb{N} \tag{2.20}$$

For $k = 0$ the statement is true. Let us then prove that if it holds for any $i \in [1, 2, ..., k-1]$ it will hold also for k. If we take the conditional expansion of relation 2.18 we have

$$VS_{t,\tau} = -\frac{2}{\tau}\mathbb{E}^{\mathbb{Q}}[\log\left(\frac{Z_{t+\tau}}{F_{t,\tau}}\right)|\mathcal{F}_t] \approx -\frac{2}{\tau}\sum_{i=0}^{k} a^k_i \frac{M^i_{t,\tau}}{F^i_{t,\tau}} \tag{2.21}$$

Which, using the induction hypothesis, allows us to express the k-th risk-neutral moment as

$$M^k_{t,\tau} \approx -\frac{2}{\tau}\frac{F^k_{t,\tau}}{a^k_k}[\frac{\tau}{2}\sum_{i=0}^{k-1} a^k_i \frac{M^i_{t,\tau}}{F^i_{t,\tau}} + VS_{t,\tau}] = -\frac{2}{\tau}\frac{F^k_{t,\tau}}{a^k_k}[\frac{\tau}{2}\sum_{i=0}^{k-1} a^k_i (1+\beta_i VS_{t,\tau}) + VS_{t,\tau}] \tag{2.22}$$

Since by equation 2.19 we have $\frac{1}{a_k^k}\sum_{i=0}^{k-1} a_i^k = -1$, we can replace the term to get relation 2.20. We finally derive the relation of the statement by using $1 + x \approx e^x$. □

Note 2.2. The reason why we have chosen the exponential instead of the original approximation is that it generally proves more convenient for well-known stochastic processes, in particular it turns to be exact when Z_t is a Brownian motion.

We remark that since the relation derived is model-free the non-structural nature of the approach is still completely preserved.
Now, given a sequence of strikes $K_1, ..., K_p$, by defining the vectors $\Delta_{t,\tau} = [\Delta_{t,\tau}^{K_1}, ..., \Delta_{t,\tau}^{K_p}]$ and $\nu_{t,\tau} = [\nu_{t,\tau}^{K_1}, ..., \nu_{t,\tau}^{K_p}]$ we can rewrite the second relation in 2.11 as

$$\Delta_{t,\tau} \approx HD, \quad \nu_{t,\tau} \approx HW \tag{2.23}$$

Where

$$D[k] = \frac{\partial F_{t,\tau}^k e^{\beta_k VS_{t,\tau}}}{\partial F_{t,\tau}} = kF_{t,\tau}^{k-1} e^{\beta_k VS_{t,\tau}}, \tag{2.24}$$

$$W[k] = \frac{\partial F_{t,\tau}^k e^{\beta_k VS_{t,\tau}}}{\partial VS_{t,\tau}} = \beta_k e^{\beta_k VS_{t,\tau}} \tag{2.25}$$

The high degree of accuracy of the method derived in this section relies on the hypothesis that the futures and the variance swap are actually the two main driving components of the underlying. This implicit statement will be significantly validated in the empirical analysis of Chapter 3. For the moment we are interested in the last essential step for the computation of the Greeks which is the derivation of the exponential coefficients. Given that the values $F_{t,\tau}$ and $VS_{t,\tau}$ are practically and, as we will see, also theoretically available, we only need to find an efficient approximation of the risk-neutral moments. This task will be completed in the following section.

2.4 Computation of the risk-neutral moments

Beside the case for $k = 1$ the risk-neutral moments are not directly observable since there are no contracts allowing to trade them. In order to extract their values we will minimize the squared distance between the actual option price and an expression similar to 2.7, derived from the orthogonal expansion. The k-th risk-neutral moments will thus correspond to the components of the related least squares coefficient vector.We will first rearrange the original relation as follows

$$V_{t,\tau}^{\Psi} \approx A_0^{\Psi} + c_1 A_1^{\Psi} + ... + c_n A_n^{\Psi} \tag{2.26}$$

Where

$$A_k^{\Psi} = \sum_{i=0}^{k} w_{k,i} \int_0^{+\infty} x^i \Psi(x)\phi(x)dx, \quad c_{t,\tau}^k = \sum_{i=1}^{k} w_{k,i} M_{t,\tau}^i \tag{2.27}$$

Given a sequence of strikes $K_1, ..., K_p$ we will then define the values

$$\mathbf{Y} = [C^{K_1}, ..., C^{K_p}, P^{K_1}, ..., P^{K_p}]' \tag{2.28}$$

$$\mathbf{X}_0 = [A_0^{K_1}, ..., A_0^{K_M}, B_0^{K_1}, ..., B_0^{K_M}]' \tag{2.29}$$

$$\mathbf{X} = \begin{bmatrix} A_1^{K_1} & ... & A_n^{K_1} \\ . & . & . \\ . & . & . \\ A_1^{K_M} & ... & A_n^{K_M} \\ B_1^{K_1} & ... & B_n^{K_1} \\ . & . & . \\ . & . & . \\ B_1^{K_M} & ... & B_n^{K_M} \end{bmatrix} \tag{2.30}$$

Given the vector $\mathbf{Y}^* = \mathbf{Y} - \mathbf{X}_0$, we will estimate the risk-neutral moments by solving the following problem

$$[\hat{c}_{t,\tau}^1, ..., \hat{c}_{t,\tau}^n] = \arg\min_{c_1,...,c_n} (\mathbf{Y}^* - \mathbf{X}c)'(\mathbf{Y}^* - \mathbf{X}c) \tag{2.31}$$

The values $\hat{c}_{t,\tau}^1, ..., \hat{c}_{t,\tau}^n$ will be called *expansion coefficients.* Since the terms $A_i^{K_j}$ may exhibit an increasing degree of multicollinearity, as n grows the columns of $\mathbf{X}$ could become linearly dependent. This means that it is not possible to apply the OLS method in this case as $\mathbf{X}'\mathbf{X}$ would be singular and thus not invertible. We will thus work out the least squares coefficient vector via a principal component regression (PCR) technique. To apply this method we will follow a sequence of steps.

1. First of all we have to center the cloud of $2M$ points of $\mathbf{Y}^*$ by choosing the suitable vector of parameters θ on which the kernel ϕ will depend upon. The centering of $\mathbf{Y}^*$ is done in order to work with a vector space instead of an affine one. We will thus have to find θ such that $\overline{\mathbf{Y} - \mathbf{X}_0(\theta)} = 0$, where the bar denotes the sample mean of the vector components. The same has to be done for the matrix $\mathbf{X}$. We will center each column vector fo $\mathbf{X}$ by subtracting the respective mean to

its components.The standardized matrix will thus be $\mathbf{X}-u[\overline{\mathbf{X}_{.,1}}, ..., \overline{\mathbf{X}_{.,n}}]$, where u is the vector of ones. Without loss of generality we will assume that $\mathbf{X}$ is already standardized. This implies that $\mathbf{X}^{'}\mathbf{X}$ will correspond to its covariance matrix.

2. Subsequently we want to find the n dimensional orthogonal vectors that most explain the variance of the row vectors of $\mathbf{X}$. These are called *weight vectors*. We will start from $\mathbf{w}_1$, which has to satisfy

$$\mathbf{w}_1 = \arg \max_{\|\mathbf{w}\|=1} \left\{\mathbf{w}^{'}\mathbf{X}^{'}\mathbf{X}\mathbf{w}\right\} = \arg \max \left\{\frac{\mathbf{w}^{'}\mathbf{X}^{'}\mathbf{X}\mathbf{w}}{\mathbf{w}^{'}\mathbf{w}}\right\} \tag{2.32}$$

This problem can be solved by setting the first order conditions for the following Lagrangian

$$\mathcal{L}(\mathbf{w}) = \mathbf{w}'\mathbf{X}^{'}\mathbf{X}\mathbf{w} - \lambda\left(\mathbf{w}'\mathbf{w} - 1\right) \tag{2.33}$$

We must have that

$$\frac{d\mathcal{L}(\mathbf{w})}{d\mathbf{w}} = 0 \Rightarrow 2\mathbf{w}^{'}\mathbf{X}^{'}\mathbf{X} - 2\lambda\mathbf{w}^{'} = 0 \Rightarrow 2\mathbf{X}^{'}\mathbf{X}\mathbf{w} - 2\lambda\mathbf{w} = 0 \Rightarrow \mathbf{X}^{'}\mathbf{X}\mathbf{w} = \lambda\mathbf{w} \tag{2.34}$$

Which implies that

$$\mathbf{w}_1 = \arg \max \left\{\frac{\mathbf{w}^{'}\mathbf{X}^{'}\mathbf{X}\mathbf{w}}{\mathbf{w}^{'}\mathbf{w}}\right\} = \max\left\{\lambda_1, ..., \lambda_n\right\} \tag{2.35}$$

With $\lambda_1, ..., \lambda_n$ real non negative eigenvalues of $\mathbf{X}^{'}\mathbf{X}$. The vector that maximizes the variance of the projections thus corresponds to the maximal eigenvector of length 1. Given the weight vectors $\mathbf{w}_1, ..., \mathbf{w}_{k-1}$, to find $\mathbf{w}_k$ we apply the same method to the matrix

$$\hat{\mathbf{X}}_k = \mathbf{X} - \sum_{s=1}^{k-1} \mathbf{X}\mathbf{w}_s\mathbf{w}_s^{'} \tag{2.36}$$

The row vectors of $\hat{\mathbf{X}}_k$ are the projections of the row vectors of $\mathbf{X}$ on the space orthogonal to the span of $\mathbf{w}_1, ..., \mathbf{w}_{k-1}$. Since the eigenvectors are orthogonal to each other and since

$$\mathbf{w}_k = \arg \max \left\{\frac{\mathbf{w}^{'}\hat{\mathbf{X}}_k^{'}\hat{\mathbf{X}}_k\mathbf{w}}{\mathbf{w}^{'}\mathbf{w}}\right\} \tag{2.37}$$

we have that the weight vector $\mathbf{w}_k$ corresponds to the normalized eigenvector with the k-th greatest eigenvalue. The orthonormal matrix $\mathbf{W} = [\mathbf{w}_1, ..., \mathbf{w}_n]$ will thus result from a spectral decomposition of $\mathbf{X}^{'}\mathbf{X}$.

3. Now consider the matrix $\mathbf{V} = \mathbf{XW}$, whose i-th column corresponds to the projection of the $2M$ row vectors of $\mathbf{X}$ on the weight vector $\mathbf{w}_i$. These projections are called *principal components*. We will only include the principal components that explain together the greatest percentage of the variance of the rows of $\mathbf{X}$. We will thus only consider the submatrix $\mathbf{V}_{.,1:s} = \mathbf{XW}_{.,1:s}$ of the first s columns. Finally, let us rearrange the original regression as follows

$$\mathbf{Y}^* = \mathbf{X}c + \varepsilon = \mathbf{XW}_{.,1:s}\mathbf{W}'_{.,1:s}c + \varepsilon = \mathbf{XW}_{.,1:s}\gamma + \varepsilon \tag{2.38}$$

Where we have used the fact that $\mathbf{W}^{-1}_{.,1:s} = \mathbf{W}'_{.,1:s}$ since the columns of $\mathbf{W}$ form an orthonormal basis. The fact that $\mathbf{W}_{.,1:s}$ is made of linearly independent vectors implies that we can find $\hat{\gamma}$ via standard OLS i.e. $\hat{\gamma} = (\mathbf{W}'_{.,1:s}\mathbf{W}_{.,1:s})^{-1}\mathbf{W}'_{.,1:s}\mathbf{Y}^*$. Since $\hat{\gamma} = \mathbf{W}'_{.,1:s}\hat{c}$, by multiplying $\mathbf{W}_{.,1:s}$ on both sides we get our result

$$\hat{c} = \mathbf{W}_{.,1:s}\hat{\gamma} \tag{2.39}$$

4. Finally, the vector of risk-neutral moments $M_{t,\tau} = M^1_{t,\tau}, ..., M^n_{t,\tau}$ is retrieved by inverting the linear equation of the second relation in 2.27 as follows

$$\hat{c} = \mathbf{T}M_{t,\tau} \Rightarrow M_{t,\tau} = \mathbf{T}^{-1}\hat{c} \tag{2.40}$$

Where $\mathbf{T}$ is the triangular matrix of orthogonal polynomial coefficients

$$\mathbf{T} = \begin{bmatrix} w_{1,1} & 0 & ... & 0 \\ w_{2,1} & w_{2,2} & ... & 0 \\ . & . & . & 0 \\ w_{n,1} & w_{n,2} & ... & w_{n,n} \end{bmatrix} \tag{2.41}$$

Since $\det \mathbf{T} = w_{1,1}...w_{n,n}$ and $w_{i,i} \neq 0$ for any $i \in [1, ...n]$, the matrix is non singular and thus invertible.

2.5 Choosing the kernel

We have always taken for granted so far the form of the kernel ϕ, an indirectly determinant function for the estimations of the previous sections and an essential element for the characterization of the non-structural nature of this method. According to Proposition 1.26, to expand the RND in the measure $\mu = \phi^{-1}$ we need ϕ to satisfy the following

i) $\int |x|^k \phi(x)dx < +\infty, \qquad \forall\, k \in \mathbb{N}$

ii) $\int (f^{\mathbb{Q}}_{t,\tau}(x))^2 \phi^{-1}(x)dx < +\infty$

iii) $\exists\, \alpha > 0 \quad s.t. \quad \phi(x) = \mathcal{O}(e^{-\alpha|x|}) \quad as \quad |x| \to +\infty$

This means that our approach is intended as non-structural since the expansion and thus the estimation of $f^{\mathbb{Q}}_{t,\tau}$ does not rely on assumptions concerning its parametric form but only on weak inferences of two general properties that are the support and the tail decay. If, for instance, we define ϕ as having support on $\mathbb{R}^+$ and exponential tail decay, to perform an orthogonal expansion we should implicitly assume that $(f^{\mathbb{Q}}_{t,\tau})^2$ is positively supported as well and that it decays at least exponentially, without any reference whatsoever to an underlying native model. Now, a typical approach in the literature is to consider the RND that describes the process $\log(\frac{Z_{t+\tau}}{Z_t})$, i.e. the standard log-returns, which would set the whole $\mathbb{R}$ as support. In our case this is not possible, though, because the method we have used to compute the Greeks and in particular equation 2.16 requires $\frac{M^k_{t,\tau}}{F^k_{t,\tau}}$ to be positive, which may not be true for some k. This restricts our support for $f^{\mathbb{Q}}_{t,\tau}$ and thus ϕ to $\mathbb{R}^+$. Considering the fixed constraints *ii*) and *iii*) as well, our suggested kernel ϕ will be the following "double-beta" density :

$$\phi(x) \sim w\left(\frac{x}{\lambda_1}\right)^{a_1}\left(1-\frac{x}{\lambda_1}\right)^{b_1}\mathbb{1}_{0\leq x\leq a_1} + (1-w)\left(\frac{x}{\lambda_2}\right)^{a_2}\left(1-\frac{x}{\lambda_2}\right)^{b_2}\mathbb{1}_{0\leq x\leq a_2} \tag{2.42}$$

Where the parameters satisfy

$$a_1, b_1, a_2, b_2, \lambda_1, \lambda_2 > 0, \quad 0 \leq w \leq 1 \tag{2.43}$$

We have seen in the previous section that the vector of parameters $\theta = [a_1, ..., w]$ must be chosen before estimating the risk-neutral moments and in such a way that the mean of the difference between the option prices and the A_0 coefficients is set to zero.
The advantages of this choice for the kernel rely upon the efficiency of the double-beta in fitting the first expression of 2.9 even for relatively low expansion orders n. Additionally, the bounded positive support solves at the same time the uncomfortable requirements of an asymptotic tail decay for ϕ and of the uncertain existence of the risk-neutral moments, now guaranteed by the continuity of ϕ on a compact domain.The choice of a bounded support solves another potential problem concerning our expansion. As extensively treated in [8], the risk-neutral moments $M^k_{t,\tau}$, with $k > 1$, could explode for finite t, which is in particular true for processes where the tail decay is slower

than the lognormal density one. This would imply the invalidation of relation 2.9, unless the risk-neutral moments of the sequence of pseudo-densities $f^n_{t,\tau}$ is always kept finite, which is exactly the case for our bounded support kernel. Clearly, some questions on the accuracy of the expansion may arise also observing that a bounded positive support does not include the case for a RND strictly positive on the entire $\mathbb{R}^+$. Nevertheless, as well outlined in the numerical illustration of [9] , even in this scenario and with possible explosive moments, by adjusting the arbitrarily large support, the efficiency of the approximation of $f^{\mathbb{Q}}_t$ and of the Greeks is not compromised and the method perfectly captures the form of the RND predicted by an *a priori* known model.

3

Empirical results

In this section we will first test the solidity of our conjectures on the firm backbone of real data and then see the computationally implemented functions at work. We will start by delving into the details concerning the SPX and the CBOE VIX, the two fundamental indexes we will work with. Our primary aim will be to remark a functional link between the VIX and the variance swap. Thereafter we will perform a granular analysis on a vast panel of options to find the risk factors and prove the hypotheses after which section 2.3 and 2.4 were elaborated. The final section will be dedicated to a study of the shape of the RND and of the Greeks plots as the market states change affecting their structures. The work will be completed with an explanation of the hedging gain measurement and of the possible solutions to a factors correlation problem emerging in low volatility periods.

3.1 The SPX and the VIX : a brief retrospective

In section 2.3 we developed our methodology for the computation of the Greeks after the primary hypothesis that the major determinants for the behaviour of the option could be addressed to the futures and the variance swap on underlying. This assumption lies at the heart of our hedging approach and it will be now provided with strong empirical support, studying the benchmark case of SPX vanilla options. Our aim is to extract the two main contributors to the variance of the option and highlight their direct link

with the aformentioned tradable contracts.
Let us start with introducing the two stock market indexes guiding our analysis. The first is the very SP500 or SPX, the most referenced index for the description of the US equity market, corresponding to a free-float adjusted capitalization-weighted sum of the 500 largest listed companies in the US market, selected and measured in terms of eligibility criteria such as market capitalization, liquidity and public float. Denoting with P_i the price of the i-th stock and with Q_i the number of publicly available shares, thus excluding those held by insiders and non public shareholders, the formula for the calculation is

$$SPX = \frac{\sum_i P_i Q_i}{divisor} \tag{3.1}$$

where the divisor is a proprietary term adjusted by the rating agency to prevent that that any change to the stocks while holding the stock prices constant may alter the market value. Multiplying and dividing the right member by the numerator we see that the index is the sum of the market capitalization weights of each stock times the total market value. The impact of the stock on the index is thus proportional to its market capitalization.
The second index is the VIX, a computationally more elusive term tracking the SPX volatility. To understand the rationale behind its formulation we should start from its native stochastic volatility model, where the dynamics of the asset price S_t with variance V_t in the risk-neutral measure $\mathbb{Q}$ are dictated by the following equations

$$d\ln S_t = \left[r - q - \frac{V_t}{2} + \lambda\left(\mu_J - e^{\mu_J + \frac{\sigma_J^2}{2}}\right)\right]dt + \sqrt{V_t}dW_t + J_t dN_t - \lambda\mu_J dt \tag{3.2}$$

$$dV_t = -\Delta_v \theta dt + vV_t^{\gamma} dB_t^* \tag{3.3}$$

In this model W_t and B_t are two correlated Wiener processes, N_t is a Poisson process with rate λ and indipendent from W_t and B_t, J_t is an indipendent normal with mean μ_J and variance σ_J, r,q and Δ_V are respectively the risk-free rate, the dividend yield and the volatility risk premium and the remaining parameters are set to include fundamental models such as the Heston's or Hull and White's ones which follow from the case $\lambda = 0$ and $\gamma = 1$ or $\gamma = 1/2$. To outline our central identity we will now provide some relations whose technical derivation is specified in [10] and will be here omitted for the sake of an intuitive exposition. We start by dividing equation 3.2 by dt, integrate both members from t to $t + \tau$ and take the risk-neutral expected value to obtain

$$E_t^{\mathbb{Q}}\left(\ln\frac{S_{t+\tau}}{S_t}\right) = \left[r - q - \lambda\left(e^{\mu_J + \frac{\sigma_J^2}{2}} - (\mu_J + 1)\right)\right]\tau - \frac{1}{2}\int_t^{t+\tau} E_t^{\mathbb{Q}}(V_s)ds \quad (3.4)$$

The last integral is the risk-neutral expected cumulative variance of the asset and its square root discounted at the present value is the quantity that the VIX aims to track. To estimate it we just have to find a way to compute the left member of the equation. This is done by replicating the term $\ln\frac{S_{t+\tau}}{S_t}$ via a portfolio of European options which, with a continuous stream of strike prices, is given at expiration time $t+\tau$ and with strike K_0 by

$$\Pi_{t+\tau}(K_0, t+\tau) = \int_0^{K_0} \frac{P_{t+\tau,\tau}(K)}{K^2}dK + \int_{K_0}^{\infty} \frac{C_{t+\tau,\tau}(K)}{K^2}dK \quad (3.5)$$

Where $P_{t+\tau,\tau}(K)$ and $C_{t+\tau,\tau}(K)$ are the put and call options expiring at $t+\tau$ with strike price K. This identity can then be expanded by using the Carr-Madan formula elaborated in [11] to get

$$\Pi_{t+\tau}(K_0, t+\tau) = \frac{S_{t+\tau} - K_0}{K_0} - \ln\frac{S_t}{K_0} - \ln\frac{S_{t+\tau}}{S_t} \quad (3.6)$$

This relation can the be discounted at time t by taking the risk-neutral expectation on both sides

$$e^{r\tau}\Pi_t(K_0, t+\tau) = \frac{F_{t,\tau} - K_0}{K_0} - \ln\frac{S_t}{K_0} - E_t^{\mathbb{Q}}\left[\ln\left(\frac{S_{t+\tau}}{S_t}\right)\right] \quad (3.7)$$

with $F_{t,\tau}$ forward price at time t with maturity at $t+\tau$. If we compute this value at $K_0 = F_{t,\tau}$, recalling the definition of variance swap in section 2.3, we see that

$$e^{r\tau}\Pi_t(F_{t,\tau}, t+\tau) = -E_t^{\mathbb{Q}}\left[\ln\left(\frac{S_{t+\tau}}{F_{t,\tau}}\right)\right] = \frac{\tau}{2}VS_{t,\tau} \quad (3.8)$$

We can now present the definition of VIX, taken directly from the CBOE White Paper:

$$\mathbf{VIX}_t^2 10^4 := \frac{2}{\tau}\sum_i \frac{\Delta K_i}{K_i^2} e^{r\tau} Q(K_i) - \frac{1}{\tau}\left[\frac{F_{t,\tau}}{K_0} - 1\right]^2 \quad (3.9)$$

In this equation τ is the time to expiration, set in such a way that the 30-day expected SPX volatility is measured and $Q(K_i)$ is the bid-ask average price of the options that are call for $K_i > K_0$ and put for $K_i < K_0$, where K_0 is the first strike below the forward $F_{t,\tau}$. Since the first term corresponds to the discrete version of $e^{r\tau}\Pi_t(F_{t,\tau}, t+\tau)$ we can rewrite the definition as

$$\mathbf{VIX}_t^2 10^4 = \frac{2}{\tau}e^{r\tau}\Pi_t(F_{t,\tau}, t+\tau) - \frac{1}{\tau}\left[\frac{F_{t,\tau}}{K_0} - 1\right]^2 = VS_{t,\tau} - \frac{1}{\tau}\left[\frac{F_{t,\tau}}{K_0} - 1\right]^2 \quad (3.10)$$

Which without the adjustment term leads us to the central identity that links the VIX to the variance swap

$$\mathbf{VIX}_t^2 10^4 \approx VS_{t,\tau} \quad (3.11)$$

This fundamental bond will be used to complete our empirical analysis in the following section and it is an essential constituent of our hedging strategy relating the second moment of the SPX to a tradable contract. Now to see how the cumulative volatility is tracked all we need to do is replace the risk-neutral expected value term in 3.7 with its expression in 3.4 so that, ignoring the adjusting terms, we can rewrite the VIX as follows

$$\mathbf{VIX}_t \approx 10^{-2}\left(\int_t^{t+\tau} E_t^{\mathbb{Q}}(V_s)ds\right)^{1/2} \quad (3.12)$$

After having delivered the essential preliminary context, we are now ready proceed with our analysis.

3.2 Analysis of driving risk factors for SPX options

In order to underpin the elemental assumption of section 2.3, we want to show that the leading drivers for SPX options are 2 and that they are functionally tied to the futures and the variance swap form. To achieve this, we perform a principal component analysis (PCA) over a panel of 1-month maturity SPX call options taken from the OptionMetrics platform. The call prices are set to the bid and ask mid value and they are categorized in a pool of 9 moneyness levels. This is done to assess how the risk drivers affect the option differently according to its intrinsic value and to lay down an empirical basis for the inference of the risk factors candidates. The spectre of intervals classifies the instrument moneyness, defined as the ratio $\frac{K_i}{S_t}$ and it gradually spans from in-the-money (ITM) to out-the-money (OTM) options with the aim of capturing the stochastic processes differences. Each interval displays a sequence of 30-day call mid prices collected at monthly frequency from 2006 to the first 4 months of 2016, for a total of 124 observations per time series. The various plots of the call prices for each moneyness scope are reported in **Fig. 3.1** with the relative range.

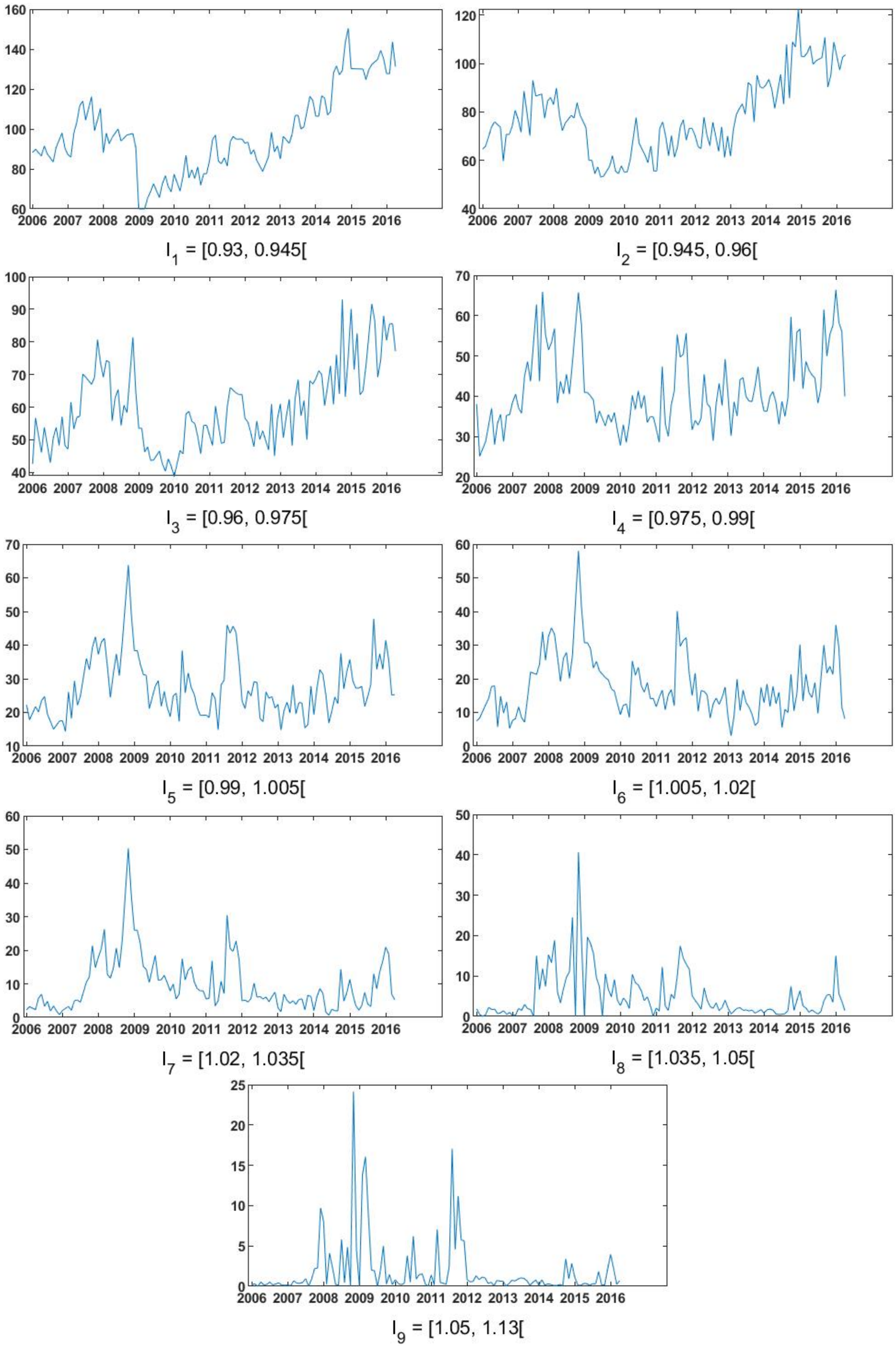


Fig. 3.1: Call mid prices from January 2006 to April 2016 classified by moneyness interval.

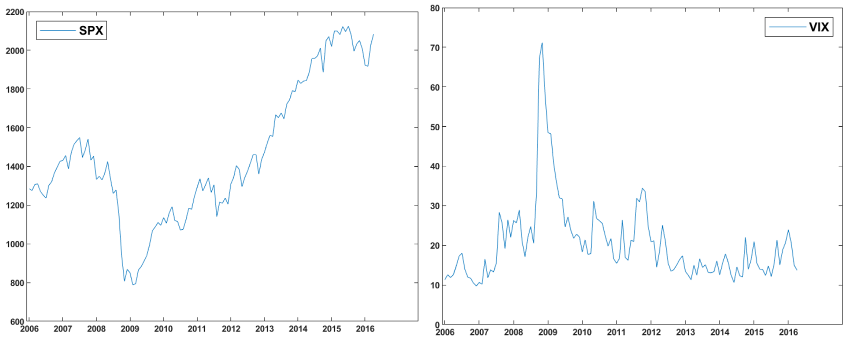


Fig. 3.2: SPX and VIX index time series from January 2006 to April 2016.

We notice how the behaviours exhibited by the time series are remarkably dissimilar and, as we move from the first to the last extremity, we see that a smooth evolution of the stochastic process unfolds, shifting from a non-stationary pattern to a prominently stationary one. This piece of evidence provides us with a line to follow for the identification of the risk factors, hinting that a linear combination of a couple of time series from closely related indexes with analogous stationary and non-stationary dynamics could be responsible for this change in trend. By looking at the very underlying, the SPX, and at the index tracking its 30-day risk-neutral expected volatility,the VIX, whose plots are displayed in **Fig. 3.2** , we witness this exact distinction of processes. The SPX parallels the patterns of ITM options in the first moneyness range as the VIX does with the OTM ones at the last intervals supposedly contributing to the variance with opposite sign coefficients. The final intuition that completes our argument is that the ITM option price as its strike goes to zero should be equal to the futures on the SPX and, with a similar reasoning, the value of the OTM instrument depends on its possibility to profit by converting to an ITM moneyness and thus to the volatility of the underlying. This conjecture is here circumscribed to SPX options due to the availability of a volatility index but it can be extended to vanilla options on stocks in general, in particular to options with the VIX as the underlying. We will test the correlation by considering the VIX^2 instead of the VIX because, by exploiting relation 3.11, we want to remark that the second risk factor is functionally intertwined with the variance swap in the same way that the SPX is linked with the futures.
To verify this hypothesis we adopt the PCA technique, a procedure already treated in section 2.4 for the computation of the risk-neutral moments via the principal component regression. The steps to follow are those already covered. First we form the 124×9 matrix $\mathbf{X}$ of call mid prices. This matrix corresponds to a cloud of 124 dots in the 9-dimensional vector space. The

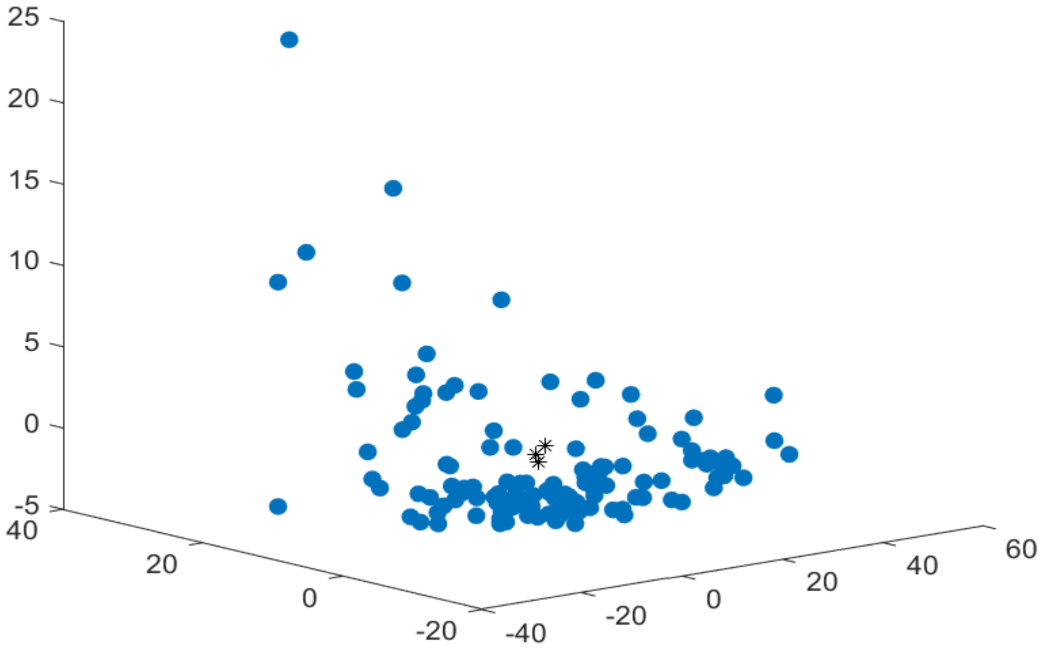


Fig. 3.3: Cloud of 124 dots centered around the origin, for the simplified case of 3 intervals of moneyness each one corresponding to a dimensional axis. The asterisks denote the heads of the orthonormal weight vectors for this dataset.

dots are then centered by removing their sample mean, which results in the matrix $\mathbf{Z} = \mathbf{X} - \bar{\mathbf{X}}$. Then the weight vectors, i.e. the orthonormal vectors maximizing the projection of the points, are computed by taking the eigenvectors of the correlation matrix $\mathbf{Z}'\mathbf{Z}$. Finally the principal components are selected. These are the projections of the dots on the weight vectors with the greatest eigenvalues i.e. the vectors explaining most of variance. A spatial representation is given in **Fig. 3.3** for the simplified case of 3 intervals of moneyness. The first confirmation of our hypothesis comes after observing the two principal weight vectors of the PCA, reported with the eigenvalues and the cumulative variance in table **1**. We see that the variance explained by the first two vectors $\mathbf{W}_1$ and $\mathbf{W}_2$, equal to the sum of their eigenvalues, is 98% of the total one meaning that almost the whole influence on the option is covered by the two main risk factors. Then, to verify that the principal components are highly matched by the SPX and the VIX^2 we compute their respective sample correlation coefficients resulting in the following values

$$\hat{\rho}_1 = corr(\mathbf{SPX}, \mathbf{ZW}_1) = 83.6824\% \tag{3.13}$$

$$\hat{\rho}_2 = corr(\mathbf{VIX}^2, \mathbf{ZW}_2) = 80.3048\% \tag{3.14}$$

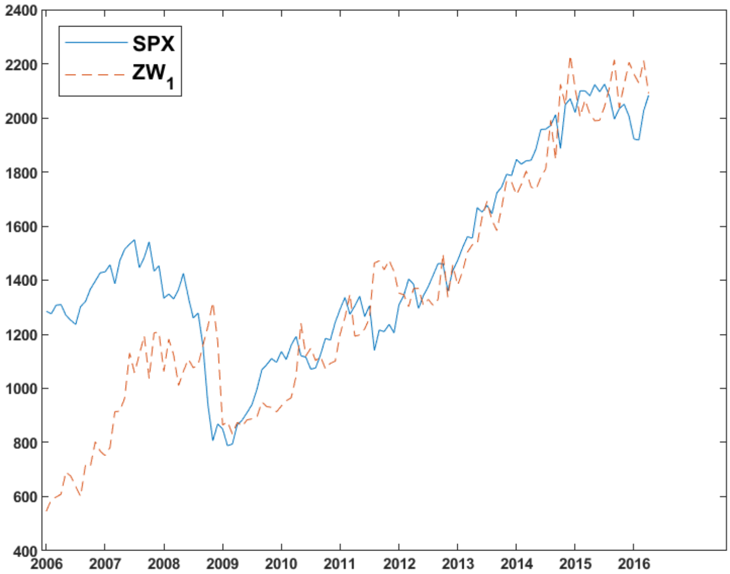
PCs	λ_k	**Cumulative**	$\mathbf{I_k}$	$\mathbf{W_1}$	$\mathbf{W_2}$	$\mathbf{R^2_{IDX}}$	$\mathbf{R^2_{PCA}}$
1	648.8e3	0.9287	0.93-0.945	0.6723	-0.2979	0.855	0.989
2	370.8e2	0.9818	0.945-0.96	0.5251	-0.1764	0.829	0.973
3	3603	0.9870	0.96-0.975	0.4037	0.1386	0.733	0.945
4	3104.3	0.9914	0.975-0.99	0.2622	0.3137	0.58	0.917
5	2032.5	0.9943	0.99-1.005	0.1641	0.4545	0.524	0.927
6	1698	0.9967	1.005-1.02	0.1018	0.4842	0.523	0.912
7	980.73	0.9982	1.02-1.035	0.0495	0.4580	0.67	0.923
8	752.86	0.9992	1.035-1.05	0.023	0.2957	0.391	0.709
9	502.33	1	1.05-1.13	0.0084	0.1546	0.352	0.537

Table 1: Elements of the principal component analysis derived from the dataset.The components $\mathbf{ZW_k}$ are sorted in descending order according to their eigenvalues. The third column lists the amount of option variance explained by the first k eigenvectors . The components of the first two weight vectors $\mathbf{W_1}$ and $\mathbf{W_2}$ are reported in the fifth and sixth columns. Finally, the last two columns display the R-square coefficients of the regression of the call prices in the interval of moneyness, respectively on the SPX and VIX observable indexes and on the two principal components.

The high sample correlations obtained, finally confirm our intuition and, via the identity 3.11, the functional link between the options and the variance swap, as well as with the futures, has been empirically established. The correlations are well described in **Fig. 3.4** where the indexes and the principal components from 2006 to 2016 are overlaid.To assess the overall linear contribution of the two weight vectors in the specific moneyness interval, in table **1** we report additionally the R^2 respectively derived by regressing in each range the call price time series first on the indexes, thus with the model $C_{t,\tau} = \beta_S SPX + \beta_V VIX^2$, and then on the weight vectors, thus following the model $C_{t,\tau} = \beta_1 W_1 + \beta_2 W_2$.

Observe that, as previously argued, the weight components corresponding to the coefficients of the linear combination are significantly positive for the SPX and negative or close to zero for the VIX^2 in the ITM range and almost specularly inverted for the OTM region. Notice that the ATM options trend is addressed almost solely to the second principal component thus indicating the VIX^2 as a close unique determinant for the central intervals.

The second observation we make concerns the dynamics of the VIX^2 with respect to the principal component from 2003 to 2016. We notice that the



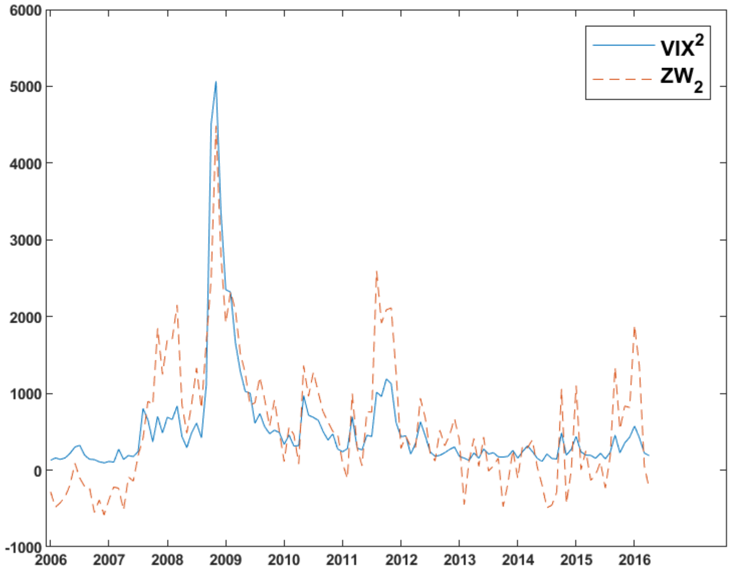


Fig. 3.4: SPX over first principal component time series and VIX^2 over second principal component time series.

two time series meet an increasing divergence thus losing their correlation as we progress through the last years. This gradual detachment of the VIX^2 statistically indicates that a factor with non null correlation emerges more and more relevantly contaminating the purely orthogonal component. We can indeed argue that from 2006 to 2013 the VIX^2 roughly parallel its orthogonalized form $\mathbf{VIX}^2 - \frac{Cov(\mathbf{SPX},\mathbf{VIX}^2)}{Var(\mathbf{SPX})}$, completely uncorrelated with the SPX, and that thereafter, as it soars from the second principal components, it slowly incorporates the additional component exhibiting an ascending negative correlation. Indexing the 30-days expected volatility of the SPX and thus measuring the level of uncertainty in the market it is curiously unusual for the VIX to report historically low values in the 3 years, distinctly incompatible with the erratic pattern of the second component. This phenomenon is known as VIX puzzle and, among the vast literature it originated, it is perhaps arguably economically ascribable to the more typically recognized leverage effect. This is the generally negative correlation observed between an asset return and its volatility. More specifically, the recurring tendencies are asymmetric : stock prices declined are paralleled by more than proportional decreases in volatility, in opposition to the steadier volatility descents as asset prices rise, this latter circumstance being our case. The term *leverage* carries the sign of the predominant cause interpretation, suggesting that the rise in asset prices decreases the leverage level relative to the equity, resulting in less riskier and thus less volatile stocks. From a wider perspective this could explain the SPX-VIX negative correlation assuming that the 500 companies took part to an analogous mechanism on an aggregate scale.

As far as out technical analysis is concerned, this spike of correlation comes as a detrimental occurrence for the hedging strategy, which is undermined wheneverthe VIX loses a faithful measure of the actual market turbulence. For the hedging to work efficiently, indeed, we require the portfolio to be resilient to risk i.e. we look for the condition $\frac{P_t}{\xi_{t,\tau}} = 0$. This could only be true if the mixed partial derivatives of the risk factors are zero, thus if they are reciprocally constant and a necessary condition for this is that they have null covariance. This issue could be circumvented via several alternatives, such as adopting an indirect form of hedging instead of directly adjusting the Greeks in accordance with the volatility. While some of the methods to soothe the leverage effect will be mentioned in the last section, for the moment we want to highlight the possibility, suggested by our results, of adopting ATM options. As already showed, the table indicates that the variance of ATM call options depends almost exclusively on the second principal component, which means that these instruments are much less sensitive to the leverage effect than those belonging to the other moneyness intervals. The reason why they could overcome the correlation problem is that they tend to require almost only one Greek, the vega, to synthesize the hedging portfolio which clearly obviates the mixed partial derivatives problem previously mentioned. Additionally this is only true for the ATM strip because even if the weights of the SPX components for the last OTM ranges are almost zero this is true for the VIX^2 component as well, indicating that the two are very far from being variance drivers as influential as in the ATM interval.

3.3 Analysis of the risk-neutral moments dependency from the variance swap

We now move on to empirically sustain the second assumption, which is the relation 2.17 of section 2.3. This is another key approximation since it defines the form of the Greeks and the quality of the hedging is thus directly dictated by its accuracy. Beside its already presented theoretical derivation, we want to evaluate the validity of the heuristic equation by taking the logarithm on both sides and rearranging the terms to perform the following regression

$$Y_t = \beta_k X_t + \epsilon_t^k \quad k = 2, 3, 4, 5 \tag{3.15}$$

with

$$Y_t := \log(M_{t,\tau}^k) - k \log(F_{t,\tau}), \quad X_t = VS_{t,\tau} \tag{3.16}$$

Similarly to the previous section, the dataset will be a time series of call and put options with monthly maturity, thus $\tau = 30$ days, collected at monthly frequency for a total span of 41 values. The risk-neutral moments $M_{t,\tau}^k$ are computed following the PCR procedure explained in section 2.3 with expansion order n set to 5. The variance swap in the regression is instead computed after the estimation of the RND via the fundamental relation 2.9. We compute the variance swap in the regressor X_t by adopting our method, replacing the call option payoff $(x - K)^+$ with $\log\left(\frac{x}{F_{t,\tau}}\right)^{-\frac{2}{\tau}}$ for $\Psi(x)$ in the second identity, which is the expression of H_k^Ψ. The regressions are made for $k = 2, 3, 4$ and 5 and the respective R-square values are reported in the table below, together with the estimated OLS coefficients $\hat{\boldsymbol{\beta}}_k$, the actual dependancy factors of 2.16 here identified with their time average as $\boldsymbol{\beta}_k = \frac{1}{N}\sum_{t=1}^N \frac{1}{VS_{t,\tau}} \log\left(\frac{M_{t,\tau}^k}{F_{t,\tau}^k}\right)$ and the p-values.

k	$\mathbf{R^2}$	$\hat{\boldsymbol{\beta}}_\mathbf{k}$	$\boldsymbol{\beta}_k$	**p-value**
2	0.94948	0.00276	0.0028	6.7749e-27
3	0.94779	0.008035	0.00821	1.2904e-26
4	0.94495	0.015605	0.01606	3.6261e-26
5	0.94106	0.025284	0.02623	1.3768e-25

The degree of efficiency of our approximation is optimal regardless of the moment order, as eloquently displayed by the peaking measurements of the R-square values. The precision of this crucial functional relation is further confirmed by the closeness of the $\hat{\boldsymbol{\beta}}_k$ and the $\boldsymbol{\beta}_k$ terms, almost identically matched despite computationally independent. The adherence of the left to the right member is then ultimately exposed in **Fig. 3.5**, where the risk-neutral moments time series and their heuristic approximations are being overlaid.

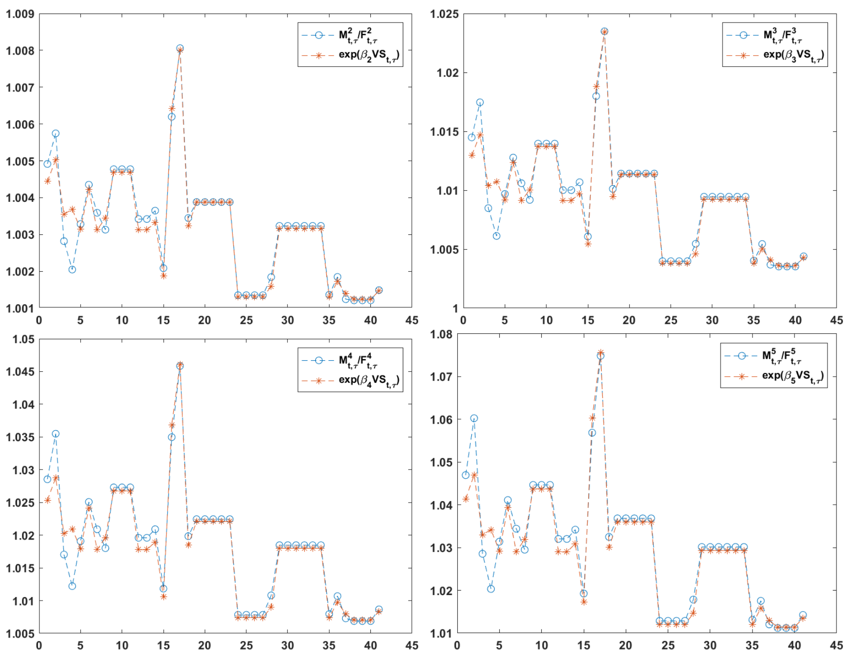


Fig. 3.5: Time series of $M^k_{t,\tau}/F^k_{t,\tau}$ and $e^{\hat{\beta}_k VS_{t,\tau}}$ overlaid.

3.4 How to read the RND in turbulent times : from the 2008 crisis to COVID-19

We can now finally enter in the main section, where, over the firm building blocks of our verified assumptions, we will put hand to a consistent set of results concerning the RND form in different circumstances.
This paragraph is formally complementary to the subsequent one, since the cross sectional approach here presented to show how information could be derived from this sophisticated market snapshot will be followed by a time series analysis aimed at illustrating how to measure the hedging efficiency in extensive trading periods. In particular, this section serves the double purpose of breaking the ice with the interpretation of the plots extracted and of casting a light on the manifold applications of the RND which, far from being employed merely for hedging or option pricing practices, could range its scope from micro to macroeconomic enviroments. Its diverse functionalities have originated a prolific varied literature in the last decade as this distribution offers a precious prism, filtering a stream of information on the risk preferences and the expectations of the traders. Avoiding to pigeonhole our tool as a mere Greeks estimator and promoting an intelligent, aware and eclectic employment of its properties, we briefly list here some of the many services and insights it could provide:

- **Analysis of the market average beliefs revealed by explicit trends in the density**. This is done by studying the center of the distribution and its change over a periodic span. It could be used by the investor to adjust trading positions and adopt VAR-wise development of confidence intervals for movement predictions.

- **Forecasting of financial crises after the analysis of extreme shocks in the third and fourth implied moments**. See [12], where the authors provide a statistical summary of the 2008 crisis with a detailed narration the SPX implied RND changes of flattening coefficients and a study of the relations of its moments and quantiles with the forwards.

- **Calibration, evaluation and monitoring of monetary policies by the central banks and the regulatory authorities**. See the ECB working papers [13] and [14], where in the latter the changes of implied moments of the RND extracted from FX options are studied around times of monetary policy decisions to gauge information on the potential predictability of exchange rates via density shape analysis and on the reaction of the market to the policy outcomes.

- **Estimation of option-implied risk aversion**. The degree of risk aversion for the representative investor determines the risk premium implicit in the option prices and tracking the average agent profile is a pivotal resource for professional practitioners and institutions to anticipate the market movements.As explained in [15],the estimation is usually performed by resorting to the historical series of the underlying asset to derive the physical density and then by confronting it with the options implied RND via their ratio, known as *pricing kernel*, which is equivalent to the utility marginal rate of substitution, thus allowing to infer its form.

There are several determinants behind the high quality of this instantaneous market barometer. Most of all, the RND collects information from multiple available options thus incorporating a spectre of strikes, unlike some frequently referenced indicators such as the futures rates, the at-the-money volatility or the risk reversal, all relying upon a very restricted set of options. Indeed, tracking the change in the futures, which for the SP500 case could be done by following the corresponding E-mini contracts, endows us with a summary potentially unbiased overview of the density mean movements which does not account for the possible presence of modes and devoid of any information on the type of transformation undergone, be it a shape-preserving traslation or a change in tail heaviness. The most likely outcome for the futures could be very different from the one suggested by the mean if its path diverges from the modal one, as showed in [13]. The at-the-money volatility is the σ of the underlying log returns required by the Black-Scholes formula to equate the option price. This measures typically varies with the strike and the maturity, depicting the so called *volatility smile* in the implied volatility surface (IVS) $\sigma_{imp}(K, T)$ and it is very narrow when compared to the variety of data mirrored in the RND. Finally, the risk reversal attempts to measure the volatility smile slope by computing the absolute difference between the implied volatilities of a call and a put options, selected to be equally out-of-the-money. Now, the steepness of the smile is the extent to which the IVS distances itself from a flat surface which depends on how the RND of log returns is far from being normal as assumed in the Black-Scholes model, a distance more faithfully measured by the distribution skewness i.e. its third standardized moment.
Without resourcing to overly crafted techniques, that would lie outside the interest of our work, we provide the main statistics features to consider when interpreting an RND plot, always aware of its intrisic difference with the real world density :

1. *Mean and standard deviation.* The first two risk-neutral moments carry

information on the market expectation on the future underlying value and on the degree of related certainty. Since the terms are always conditional on the present available knowledge, nested in the filtration $\mathcal{F}_t$, we usually observe that the RND volatility depends on the maturity of the options it is extracted from as displayed in **Fig. 3.6**. The figures report the RND of the same options after 26 days, with 31 and 5 days to maturity. The closer the traders are to the expiration date the more certain they are on the future outcome, the narrower the density. This is also the reason why we denote it as $f^{\mathbb{Q}}_{t,\tau}$, including the horizon τ in the definition. This additionally suggests that the more ebullient the times and thus the larger the uncertainty, the more platykurtic the distribution, as visible in **Fig. 3.8**. The pictures show the RND extracted from 30-days maturity VIX options via a Generalized inverse Gaussian kernel density first in a normal period and then in the tumultuous COVID-19 alarm diffusion one, where as showed in **Fig. 3.7**, the VIX suddenly escalated to outrageous levels.

2. *Asymmetry measures : skewness, (mean-mode)/standard deviation, (mean-median)/standard deviation.* Asymmetry in the distribution ascribes to the perception that the underlying has more chance to exceed (or fall behind) the most anticipated value, than instead falling behind it (or exceeding it), as it is evident for the VIX densities of **Fig. 3.8**. Furthermore a rapid truncation in one of the tail, as slightly outlined for the right tail in the first plot of **Fig. 3.6**, could mean that the market has set a limit in the believed decline or rise of the asset. Besides, as already mentioned, skewness could be regarded as a measure of the detachment from the Black-Scholes assumptions of log-normality for the returns and Brownian motion for the underlying stochastic process. To track dynamically the symmetry fluctuations, a good procedure could be the study of the pattern of the quantiles and their correlation with the market indices, as performed in [12].

3. *Kurtosis.* The standardized fourth risk-neutral moment reflects the fatness of the tails, thus revealing the risk of extreme movements in the underlying, an indicator that could be used to evaluate the market anticipation of a financial shock, as showed in [12] and **Fig. 3.9** for the 2008 financial crisis case.

4. *Modes.* The modes of a continuous distribution are its local maxima i.e. the set of most likely values. A multimodal density, such as the one presented in the first plot of **Fig. 3.9**, describes an heterogeneity of beliefs and thus a scattered probability, where the expectations of the

agents are highly dispersed around the main value. On the contrary, a unimodal density could indicate a homogeneous convergence of the traders on the shape of the underlying future distribution resulting either in a highly leptokurtic RND, such as the second of **Fig. 3.6**, or a more platykurtic one, such as the second of **Fig. 3.9**.

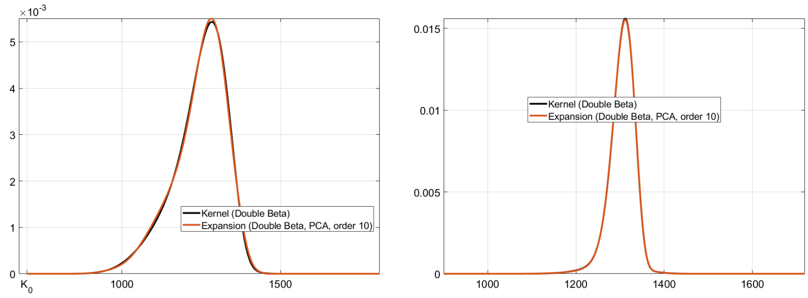


Fig. 3.6: Risk-neutral density derived via double-beta density from the same panel of SPX options on 2008-07-16 and 2008-08-11, with respectively 31 and 5 days to expiration.

Let us now draw some parallel considerations on the Greeks. On **Fig. 3.10** the Greeks for 2016 SPX options are plotted with respect to the strike and on **Fig. 3.11** we have the graphs of the deltas of the same 2016 and 2020 VIX options of **Fig. 3.8** . What we see is a concise and compact supplementary confirmation of the assumption studied in section 2.3. Recalling the results of the PCA, in particular the components of the weight vectors, we should expect this scenario :

- *ITM range.* Very high sensitivity to the underlying (delta), usually greater than 0.5, and very low sensitivity to the change in volatility (vega).
- *ATM range.* Medium or negligible sensitivity to the underlying, usually around 0.5, and high sensitivity to the change in volatility.
- *OTM range.* Very low sensitivity to the underlying and almost all sensitivity loaded on the volatility. Null sensitivity to both for extreme strikes where the option is illiquid.

The plots of **Fig. 3.10** sustain our view. The delta descends while the vega has a slightly asymmetrical bell curve form. The closer we get to the end of the RND support, the less traded are the OTM options and thus the less sensitive they are to both risk factors. In **Fig. 3.11** the effects of peaking VIX during the pandemic risk are evident when compared to the 2016 period

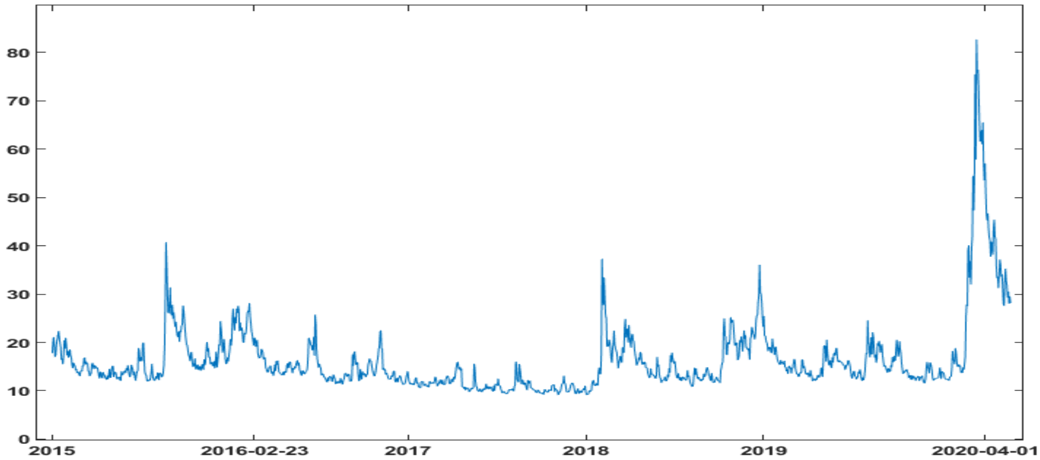


Fig. 3.7: VIX time series from January 2015 to May 2020. The dates of the panel of options selected are marked.

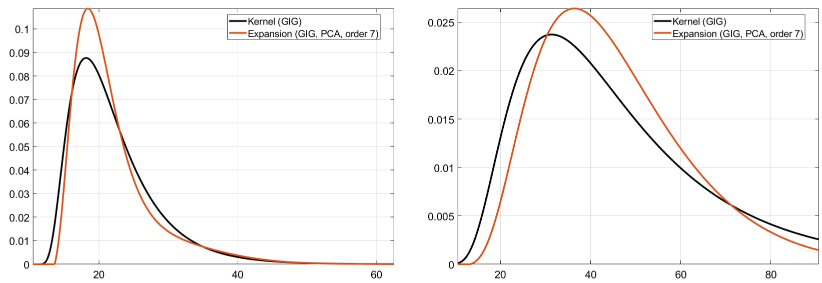


Fig. 3.8: Risk-neutral density derived via a Generalized inverse Gaussian kernel from a panel of VIX options with monthly maturity, first on 2016-02-23 and then on 2020-04-01.

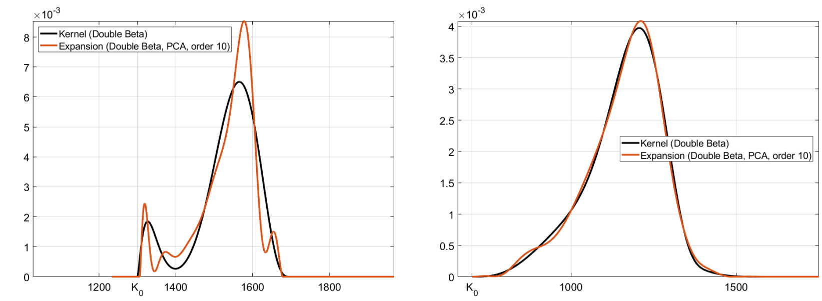


Fig. 3.9: Risk-neutral density extracted from 30-day maturity SPX options in 2007-09-19 and in 2008-09-17, when the crisis hit strongly.

of historical lows. The VIX delta curve size is much more amplified in the former case than in the latter, because the VIX is more expected to reach higher values, thus influencing even the most remote OTM options to possibly convert their state. This implies that the hedging should work optimally in 2020 as the correlation factor is low and their influence on the options is consistent meaning that the fluctuations are well captured by the direct positions taken.

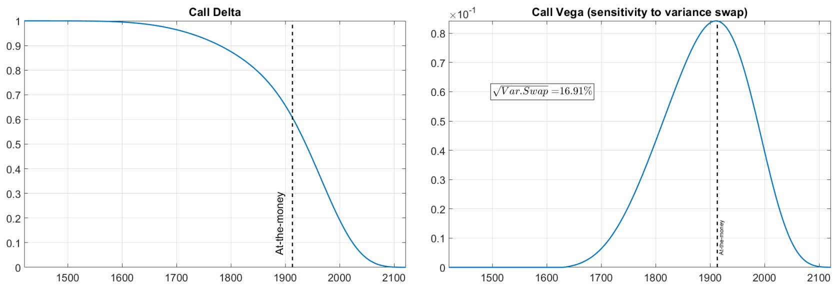


Fig. 3.10: Delta and vega of 30-day maturity SPX call options on 2016-02-23 with the strikes reported on the x-axis.

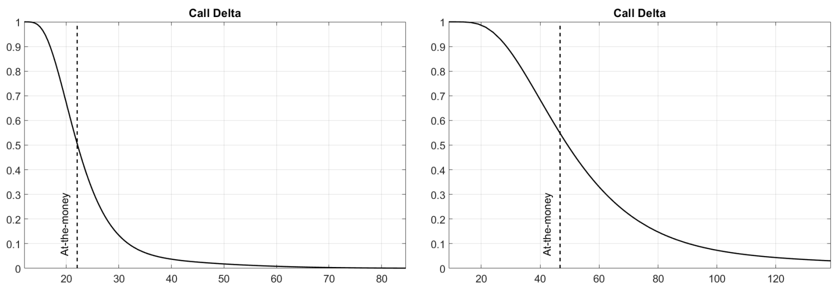


Fig. 3.11: Deltas of 30-day maturity VIX call options on 2016-02-23 and 2020-04-01.

3.5 Measuring the hedging gain and solving the correlation issues

In this final section we will illustrate how the hedging performance could be measured in a sufficiently extended time span. Afterwards, we will present some remarks on the already met technical issues emerging in low volatility regimes and we will see how to possibly solve them.
Let us first consider a set of strikes $K^1, ..., K^M$, an issuing day t_0 and an horizon τ. We will assume that the rebalancing of the portfolio weights

occurs periodically, with the time values set by $t_1 = t_0 + \frac{\tau}{N}$,...,$t_N = t_0 + \tau$. In order to lighten our notation, we will rewrite the familiar values as

$$C_i^{K_j} = C_{t_i,\tau-t_i}^{K_j}, F_i = F_{t_i,\tau-t_i}, VS_i = VS_{t_i,\tau-t_i}, \Delta_i^{K_j} = \Delta_{t_i,\tau-t_i}^{K_j}, \nu_i^{K_j} = \nu_{t_i,\tau-t_i}^{K_j} \tag{3.17}$$

The hedging is performed here on SPX options and thus we know from section 2.3 that the Greeks are taken respectively from the futures and the variance swap. The hedging is exectued on a portfolio for each strike, where the form 2.5 for the given K_j becomes

$$C_i^{K_j} + \Delta_i^{K_j} F_i + \nu_i^{K_j} VS_i + \pi_i^B B_i \tag{3.18}$$

The hedging efficiency is not measured in payoff terms but depends on the extent to which the option replication component in the portfolio is good at tracking the variation of the call price between rebalancing times. This difference , equal to $C_{i+h} - C_i$ with the strike omitted for simplicity, can be rearranged for small h as follows

$$h\frac{C_{i+h} - C_i}{h} \approx h\frac{dC_i}{dh} = h\left(\frac{\partial C_i}{\partial F_i}\frac{dF_i}{dh} + \frac{\partial C_i}{\partial VS_i}\frac{dVS_i}{dh}\right) \approx \Delta_i(F_{i+h} - F_i) + \nu_i(VS_{i+h} - VS_i) \tag{3.19}$$

Where we have used the total derivative formula and the fact that the call $C_i = C_i(F_i, VS_i)$ is mainly influenced by the two factors. The goodness of tracking can be thus framed in the *replication squared error* (RSE), which corresponds to the term

$$RSE_i^{K_j} = \left(\Delta_{i-1}^{K_j}(F_i - F_{i-1}) + \nu_{i-1}^{K_j}(VS_i - VS_{i-1}) - (C_i^{K_j} - C_{i-1}^{K_j})\right)^2 \tag{3.20}$$

In order to adapt this value to a scale-free metric and give it a meaning to its quality by confronting with another approach, we normalize it by the RSE of a benchmark hedging strategy which is the practitioners' Black-Scholes (PBS). This employs only one Greek, the delta, still equal to the sensitivity to the underlying futures but computed with the Black-Scholes closed formula which, with the asset not paying dividends, is written as

$$\Delta_i^{BS,K_j} = \frac{\partial C_i^{K_j}}{\partial F_i} = \phi\left(\frac{\ln(\frac{F_i}{K_j}) + (r + \frac{\sigma^2}{2})(T - t_i)}{\sigma\sqrt{T - t_i}}\right) \tag{3.21}$$

Where r is the risk-free interest rate and the term depends on K_j since the parameter σ is the implied volatility of $C^{K_j}i$, which is typically influenced by the strike. The respective benchmark replication squared error (BRSE) is then formulated as

$$BRSE_i^{K_j} = \left(\Delta_{i-1}^{BS,K_j}(F_i - F_{i-1}) - (C_i^{K_j} - C_{i-1}^{K_j})\right)^2 \tag{3.22}$$

Thus, given a set of M portfolios with respective strikes and N times, its average gain measuring the hedging performance is the relative value of the RSE on the BRSE, formally expressed as

$$G = \frac{1}{M}\sum_{j=1}^{M}\left(1 - \frac{\sum_{i=1}^{N} RSE_i^{K_j}}{\sum_{i=1}^{N} BRSE_i^{K_j}}\right) \tag{3.23}$$

The better the hedging strategy with respect to the PBS the closer to 1 the gain.

We have seen in section 3.2 that some troubling correlation between the SPX and the VIX, emerging in low volatility regimes, could severely negatively affect the hedging. Since, indeed, we minimize the risk by taking direct positions on both the driving factors, we should always keep an eye on their covariance which, for reasons related to overly favorable monetary policies or market conditions, could significantly soar compromising the necessary condition for the mixed partial derivatives to be null and the hedging to work. This could be viewed as the VIX no longer translating the actual risk intrinsic in the products, an occurrence which causes a distorted perception of the market movements and induces the agents to over-leverage themselves towards a sudden unpredicted crash. We should therefore expect our hedging to be optimal in the highest turmoil, where the VIX adequately responds to the havoc, and badly in the quiet times, with the PBS turning out to be more solid as the options risk is loaded more on the delta.

We briefly outline in section 3.2 some possible methods to overcome this complication. The main guiding intuition is to orthogonalize the VIX with respect to the SPX, thus working the Greek of its perpendicular component treated as the risk factor, which we wrote as $VIX - \frac{Cov(VIX,SPX)}{Var(SPX)}SPX$. This strategy is undermined by an inevitable caveat.To understand what could occur we first have to identify the possible behaviour of the historical $\mathbf{VIX}_t^2$ vector in the space of **Fig. 3.3**, where we performed the principal component analysis. We have that the coordinates of $\mathbf{SPX}_t$ and $\mathbf{VIX}_t^2$ are such that the projections over them of the centered historical cloud of dots $\mathbf{Z}_t$ are equal to the observed time series of the two indexes i.e. $SPX_t = \mathbf{Z}_t\mathbf{SPX}_t$ and $VIX_t^2 = \mathbf{Z}_t\mathbf{VIX}_t^2$. Now, since the $\mathbf{SPX}_t$ corresponds to the first weight $\mathbf{W}_1$, the vector remains still with respect to $\mathbf{W}_2, \mathbf{W}_3, ..., \mathbf{W}_n$, which is instead not true for $\mathbf{VIX}_t^2$ that, as we have seen, will lose gradually its orthogonality. A

natural adjustment consists in identifying the $\mathbf{VIX}_t^2$ orthogonal component as the new $\mathbf{W}_2$, but this is valid only if we assume that the orthogonal part will remain still with respect to the $\mathbf{W}_i$ base. This constraint could be easily violated if $\mathbf{VIX}_t^2$ departs from the space generated by $\mathbf{SPX}_t$ and $\mathbf{W}_2$, as displayed in **Fig. 3.12**, invalidating the hypothesis that it is still related to a principal factor. In order to decompose better the VIX we could extrapolate its principal components, try to relate them with some observable values and project the VIX on these as well which could be a tedious task.
Another more comfortable approach was the already discussed use of ATM options to indirectly hedge on the vega, as explained in [16]. By denoting with $C_i^{\bar{K}}$ the ATM call option we could consider the portfolio

$$C_i^K + \pi_i^{K,F} F_i + \pi_i^{K,\bar{K}} C_i^{\bar{K}} \tag{3.24}$$

with

$$\pi^{K,F} = -\Delta_i^K - \pi_i^{K,\bar{K}} \Delta_i^{\bar{K}}, \quad \pi_i^{K,\bar{K}} = -\frac{\nu_i^K}{\nu_i^{\bar{K}}} \tag{3.25}$$

By taking the F_i derivative we get

$$\Delta_i^K - \Delta_i^K - \pi_i^{K,\bar{K}} \Delta_i^{\bar{K}} + \pi_i^{K,\bar{K}} \Delta_i^{\bar{K}} = 0 \tag{3.26}$$

By taking the VS_i derivative we get

$$\nu_i^K + \pi_i^{K,F} \frac{\partial F_i}{\partial VS_i} - \nu_i^K = (-\Delta_i^K + \frac{\nu_i^K}{\nu_i^{\bar{K}}} \Delta_i^{\bar{K}}) \frac{\partial F_i}{\partial VS_i} \tag{3.27}$$

The last term can be rewritten as

$$-\frac{dC_i^K}{dF_i} \frac{dF_i}{dVS_i} - \frac{dC_i^K}{dVS_i} \frac{dVS_i}{dC_i^{\bar{K}}} \frac{dC_i^{\bar{K}}}{dF_i} \frac{dF_i}{dVS_i} \tag{3.28}$$

so that the hedging is preserved in its indirect form by including a normalized vega.

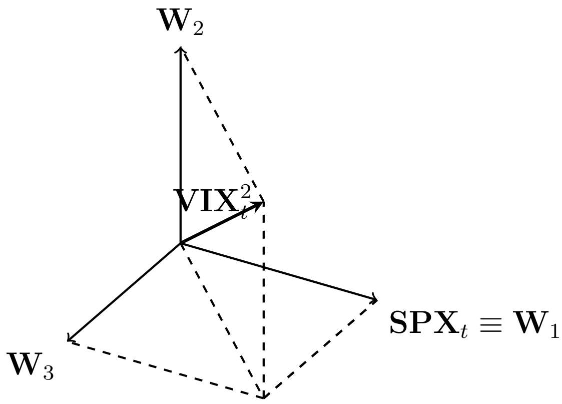


Fig. 3.12: Possible configuration for the $\mathbf{VIX}_t^2$ decomposition, with the vector departing from the weight $\mathbf{W}_2$ but leaving the space $Span(\mathbf{SPX}, \mathbf{W}_2)$ as well, guided also by the weight $\mathbf{W}_3$.

3.6 Conclusions

We have developed a non-structural approach for the estimation of the risk-neutral density from vanilla options. The derivation could be carried out under very mild regularity conditions, in particular an exponential tail decay of the kernel, that are proven in Chapter 1 after a rigorous theoretical construction. This pivotal finding helps us to generalize and refine some equivalent extraction procedures such as the Laguerre polynomials expansion or the Edgeworth expansion with Hermite polynomials. After the exposition of our strategy rationale and its meaning within the option pricing context, we empirically sustained our assumptions, above all the fact that the main risk factors determining the option variance are functionally linked with the tradable futures and variance swap contracts. This conjecture was verified via a principal component analysis on a dataset of benchmark SPX options. We have exposed the outcomes of our methodology first by providing an array of manifold applications, lying beyond the restricted field of the hedging techniques and ranging from monetary to general macroeconomic insights. We then outlined an essential handbook to interpret this instrument showing its shape alteration and those of the Greeks from tranquil to turbulent times. Finally, we described how the hedging could be measured after the time series of the replication squared errors and we addressed the difficulties emerging in low volatility periods by proposing alternative strategies, most of all the idea of changing the ratios for an indirect form of vega hedging.

A
Appendix

Theorem A.1. *Let $f(z)$ be a complex function on a nonempty connected open set $\Omega \in \mathbb{C}$ and differentiable therein. Then $f(z)$ has derivatives of all orders in Ω.*

Proof. See [1], Section 2.3. □

Note A.1. The function $f(z)$ is said to be *analytic*.

Theorem A.2 (Lebegue's Dominated Convergence Theorem)**.** *Let f_n be a sequence of complex measurable functions on $X \subseteq \mathbb{R}$ such that*

$$f(x) = \lim_{n\to+\infty} f_n(x) \tag{A.1}$$

Exists for every $x \in X$. If there is a function $g(x) \in L^1_\mu(X)$ such that

$$|f_n(x)| \leq g(x) \qquad \forall n \in \mathbb{N}, \forall x \in X \tag{A.2}$$

Then

$$f \in L^1_\mu(X), \qquad \lim_{n\to+\infty} \int_X |f_n - f| d\mu = 0 \qquad \lim_{n\to+\infty} \int_X f_n d\mu = \int_X f d\mu \tag{A.3}$$

Proof. See [2], Theorem 1.34. □

Theorem A.3. *Let $f(z)$ be a complex analytic function on a nonempty connected open set $\Omega \in \mathbb{C}$ such that $f^k(z_0) = 0$ $\forall k \in \mathbb{N}$ for some $z_0 \in \Omega$. Then $f \equiv 0$ on Ω.*

Proof. See [1], Section 3.2. □

Theorem A.4 (The Uniqueness Theorem)**.** *Let $f(x) : (a, b) \to \mathbb{C}$ with $(a, b) \subseteq \mathbb{R}$ and $f \in L^1(a, b)$. Denoting $\hat{f}(t)$ the Fourier transform of f, if we have that $\hat{f}(t) = 0$ for any $t \in \mathbb{R}$ then $f(x) = 0$ almost everywhere in (a, b).*

Proof. See [2], Theorem 9.12. □

Theorem A.5. *A complex function $f(z) = u(z)+iv(z) = u(x+iy)+iv(x+iy)$ on a nonempty connected open set $\Omega \in \mathbb{C}$ is analytic therein if and only u and v have continuous partial derivatives which verify the following Cauchy-Riemann equations :*

$$u_x = v_y, \qquad u_y = -v_x \qquad \forall x, y \in \mathbb{R} \quad s.t. \quad x + iy \in \Omega \tag{A.4}$$

Proof. See [1], Section 1.2. □

Proposition A.6 (Leibniz integral rule)**.** *Let $f(x, t)$ be a real function such that both $f(x, t)$ and its partial derivative $f_x(x, t)$ are continuous in t and x in some region of the (x, t)-plane, including $a(x) \leq t \leq b(x), x_0 \leq x \leq x_1$ for two functions $a(x)$ and $b(x)$that are both continuous and both with continuous derivatives for $x_0 \leq x \leq x_1$. Then, for $x_0 \leq x \leq x_1$*

$$\frac{d}{dx}\left(\int_{a(x)}^{b(x)} f(x,t)\,dt\right) = f\Big(x, b(x)\Big)\cdot\frac{d}{dx}b(x) - f\Big(x, a(x)\Big)\cdot\frac{d}{dx}a(x) + \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} f(x,t)\,dt \tag{A.5}$$

Proof. See [3], pp. 615-627. □

Proposition A.7 (Cauchy-Schwarz inequality)**.** *Let $(V, \langle ., . \rangle)$ be an inner product space. Then we have the following inequality*

$$|\langle v, w \rangle| \leq \|v\|\|w\| \qquad \forall v, w \in V \tag{A.6}$$

Proof. See [4], Theorem 1.35. □

Bibliography

[1] Ahlfors, L.V., *Complex Analysis, Second Edition*, McGraw-Hill, 1966.

[2] Rudin, W., *Real and Complex Analysis, Third Edition*, McGraw-Hill, 1987.

[3] Flanders, H., *The American Mathematical Monthly , Vol. 80, No. 6*, The Mathematical Association of America, 1973.

[4] Rudin, W., *Principle of Mathematical Analysis, Third Edition*, McGraw-Hill, 1976.

[5] Folland, G.B., *A Guide to Advanced Real Analysis*, The Mathematical Association of America, 2009.

[6] Shreve, S.E., *Stochastic Calculus for Finance II*, Springer Finance Textbook, 2008.

[7] Jondeau, E., Poon, S., Rockinger, M.,*Financial Modeling Under Non-Gaussian Distributions*, Springer Finance Textbook, 2007.

[8] Andersen, L.B., Poon, S., Piterbarg, V.V., *Moment explosions in stochastic volatility models*, Finance Stochastics 11 (1), 29-50, 2007.

[9] Barletta, A., Santucci de Magistris, P., Sloth, D., *It Only Takes a Few Moments to Hedge Options*, Journal of Economic Dynamics and Control 100, 251-269,2019.

[10] Duan,J.C., Chung-Ying, Y., *Jump and Volatility Risk Premiums Implied by VIX*, 2007.

[11] Carr, P., Madan, D.B.,*Option valuation using the fast Fourier transform*, Journal of Computational Finance 2 (4), 61-73,1999.

[12] Birru, J., Figlewski, S.,*Anatomy of a meltdown: The risk neutral density for the S&P 500 in the fall of 2008*, Journal of Financial Markets 15 (2), 151-180, 2012.

[13] de Vincent-Humphreys, R., Gutiérrez, J.M.P.,*A quantitative mirror on the Euribor market using implied probability density functions*, Working Paper Series, 1218, 2010.

[14] Birru, J., Figlewski, S.,*Estimating and analysing currency options implied risk-neutral density functions for the largest new EU member States*, Working Paper Series, 440, 2005.

[15] Bliss, R.R., Panigirtzoglou, N.,*Option-implied risk aversion estimates*, The Journal of Finance, 59 (1), 2004.

[16] Daluiso, R., Morini, M.,*Hedging efficiently under correlation*, Quantitative Finance, 17 (10), 1535-1547, 2017.

Summary

This thesis focuses first of all on the construction of an efficient hedging strategy for a portfolio of vanilla options, futures and variance swap contracts, where the weights of the contingent claims, known as *Greeks*, are estimated after the moments of the risk-neutral density (RND) of the underlying asset, extracted via a non-structural semi-parametric approach based on a polynomial expansion. The Greeks measure the sensitivity of the option to a determinant pilot risk factor and the ones we refer to, as well as the most extensively accounted for, are the delta and the vega, corresponding to the partial derivatives of the option price with respect to the level of the underlying and to its volatility. To accurately determine their value we need an unbiased method to extrapolate the risk-neutral density. This function, whose existence is guaranteed only in arbitrageless venues, is the probability distribution of the prices of the underlying asset in the measure $\mathbb{Q}$, which prices the contingent claims from the perspective of a risk-neutral investor. In the vast and continuously updating literature on the density extraction a first fundamental distinction is made between structural and non-structural models. The former provide a complete description of the stock prices dynamic, the latter derive instead the density by resourcing only from a partial or absent definition of the underlying stochastic process. Non-structural approaches could themselves be classified in parametric, where a direct expression of the risk-neutral density is proposed, and semi or non-parametric, where the density is estimated with approximation techniques. A widespread trading strategy is the so-called practitioners' Black-Scholes consisting in a delta hedging, with the hedging coefficient computed with the Black-Scholes formula. This is a structural approach as in the Black-Scholes setup the asset price follows a log-normal diffusion process, preserved in the risk-neutral measure $\mathbb{Q}$ for the Girsanov theorem. The main caveats of this procedure stand in the empirically proved failure of the hypotheses, since the volatility of the underlying is not constant and the process may exhibit jumps, and the absence of the vega. More and more refined models have been proposed, with a stochastic volatility and volatility of volatility, but in general any structural or non-structural parametric approach suffers from some recurring drawbacks, especially when historical data are used. When, indeed, a relatively simple model is taken the results could be biased and not adherent to the actual observations, whereas when a complex model is considered, the trade-off is detrimental for the processing time requested for the determination of the parameters. Our methodology only relies upon weak regularity conditions without imposing binding constraints for the underlying stochastic process, revisiting and generalizing some standard procedures based on polynomial expansion such as the Edgeworth series or, with the proper adaptation, the Gram-Charlier A series. The Greeks are then computed from the

derived risk-neutral moments via some cunning heuristic identities, later empirically verified with the other assumptions.
Beside this main purpose, the work is set for several additional aims. By venturing in the field of option hedging we notice how the areas of derivative pricing and economics will intertwine with our findings. We therefore often detail the whole context by focusing on the VIX puzzle phenomen explanations, outlining how the Greeks should in general change according to the option moneyness and, most importantly, refer to the many ways the risk-neutral density could be utilized in other academic areas. This is done to endow the work with a broader scope, indirectly suggesting some possible variations on the theme treated.
We begin by rigorously deriving the theoretical framework underlying the rationale of our hedging strategy. The main purpose is to formally establish the conditions under which a probability density of a stock price, itself determining the option payoff, can be expressed as an infinite linear combination of polynomials. We see that this process parallels a traditional result of Real Analysis which is the construction of a complete basis for an Hilbert space i.e., intuitively, an infinite dimensional vector space. The pivotal theorems needed to implement our methods are two. The first defines how to build the orthogonal polynomials from a recursive algorithm :

Theorem 0.1. *A sequence of orthonormal polynomials* $\{p_n(x)\}_{n=0}^{+\infty}$ *satisfies*

$$p_n(x) = A_n[xp_{n-1}(x) + B_n p_{n-1}(x) + C_n p_{n-2}(x)] \qquad n = 2, 3, ... \tag{1}$$

where

$$A_n \neq 0, \quad B_n = -\langle xp_n(x), p_n(x)\rangle, \quad C_n = -\langle xp_n(x), p_{n-1}(x)\rangle \quad n = 2, 3, ... \tag{2}$$

From which we derive the polynomials as

$$p_0(x) = w_{0,0} = 1 \tag{3}$$

$$p_1(x) = w_{1,0} + w_{1,1}x = \frac{x - \frac{\mu_1}{\mu_0}}{(\mu_2 - \frac{\mu_1^2}{\mu_0})^{1/2}} \tag{4}$$

$$p_n(x) = w_{n,0} + ... + w_{n,n}x^n = A_n[(x + B_n)\sum_{k=0}^{n-1} w_{n-1,k}x^k + C_n \sum_{k=0}^{n-2} w_{n-1,k}x^k] \tag{5}$$

where we have

$$B_n = -\sum_{k=0}^{n-1}\sum_{q=0}^{n-1} w_{n-1,k}w_{n-1,q}\mu_{k+q+1}, \qquad C_n = -\sum_{k=0}^{n-1}\sum_{q=0}^{n-2} w_{n-1,k}w_{n-2,q}\mu_{k+q+1} \tag{6}$$

With the coefficients $w_{i,j}$ derived from the computation of the non normalized terms $w'_{i,j} = A_n w_{i,j}$ and the subsequent definition of the A_n factor

$$w'_{i,j} = \begin{cases} B_n w_{n-1,0} + C_n w_{n-2,0} & \text{if } j = 0, \\ w_{n-1,j-1} + B_n w_{n-1,j} + C_n w_{n-2,j} & \text{if } j = 1, ..., n-2, \\ w_{n-1,n-2} + B_n w_{n-1,n-1} & \text{if } j = n-1, \\ w_{n-1,n-1} & \text{if } j = n, \\ 0 & \text{if } j > n. \end{cases} \tag{7}$$

$$A_n = \pm(\sum_{k=0}^{n} \sum_{q=0}^{n} w'_{n,k} w'_{n,q} \mu_{k+q})^{-1/2} \tag{8}$$

The second proposition sets the condition under which a function can be expanded as a series of orthogonal polynomials weighted by a kernel density, only required to satisfy an exponential tail decay and it is the core statement behind our approach :

Proposition 0.2. *Given a Borel measure μ on an open set $S \subseteq \mathbb{R}$ let the kernel $\phi : S \to \mathbb{R}$ and the target $f : S \to \mathbb{R}$ be two measurable functions on $(S, \mathcal{B}(S), \mu)$ with supp$(f) \subseteq$ supp$(\phi) \subseteq S$ and such that :*

i) The kernel ϕ is different from zero almost everywhere and it satisfies

$$\int |x|^k \phi^2(x) d\mu(x) < \infty, \qquad \forall\, k \in \mathbb{N} \tag{9}$$

ii) The target f belongs to the space L^2_μ, i.e.

$$\int f^2(x) d\mu(x) < \infty \tag{10}$$

Then the following holds :

1. *There exists a family of polynomials $(p_k)_{k\in\mathbb{N}}$ such that the corresponding ϕ-weighted family $(\phi p_k)_{k\in\mathbb{N}}$ is an orthonormal set in L^2_μ, i.e.*

$$\langle \phi p_k, \phi p_l \rangle_{2,\mu} = \Delta_{kl} \qquad \forall\, k, l \in \mathbb{N} \tag{11}$$

2. *The Fourier coefficients are well defined, i.e.*

$$c_k = \langle f, \phi p_k \rangle_{2,\mu} < \infty \qquad \forall\, k \in \mathbb{N} \tag{12}$$

3. *The sequence of the pseudo-densities*

$$f_n(x) = \phi(x)\sum_{k=0}^{n} c_k p_k(x) \qquad\qquad n \in \mathbb{N} \tag{13}$$

converges in the space L^2_μ.

If μ *is absolutely continuous with respect to the Lebesgue measure and if, given the Radon-Nikodym derivative* $\frac{d\mu}{dx}$*, the following condition holds :*

$$\exists\, \alpha > 0 \quad s.t. \quad \frac{d\mu}{dx}\phi^2(x) = \mathcal{O}(e^{-\alpha|x|}) \quad as \quad |x| \to +\infty \tag{14}$$

4. *The pseudo-densities* f_n *converge to the target* f *in norm, i.e.*

$$\lim_{n\to+\infty} \|f - f_n\|_{2,\mu} = 0 \tag{15}$$

Note 0.1. We recall that $f(x) = \mathcal{O}(g(x))$ as $|x| \to +\infty$ if and only if there exist two positive real numbers M and x_0 s.t. $|f(x)| \leq Mg(x)$ when $|x| \geq x_0$.

After having articulated in full the mathematical architecture behind our main results, structuring it with the respective proofs, we move to the central part where the intuition of the methodology, the financial theory and the heuristics applied to determine the pivotal hedging entities are explained and carefully derived. We start by defining formally familiar and less familiar terms of finance from the notion of arbitrage to that of self-financing portfolio and then finally lay the assumptions on which our analysis, and in general option pricing, is based upon. By denoting with Z_t the value of a generic underlying asset at a time $t \geq 0$ we set $t + \tau$ as the time of expiration for the contingent claim and state the following general assumptions :

1. The market is arbitrage-free.
2. The market is complete.
3. The market is frictionless.
4. The risk-free interest rates are set to zero without loss of generality.
5. The price of any derivative security on the underlying Z_t depends on a finite number of traded risky factors $\xi_{t,\tau} = (\xi^1_{t,\tau}, ..., \xi^q_{t,\tau})$.
6. $\exists\, n \geq 2$ $s.t.$ $\mathbb{E}^{\mathbb{P}}[Z^k_{t+\tau}] < \infty,\ k = 0, 1, ..., n.$

This sets of assumptions guarantees the existence of a unique risk-neutral probability measure, which is the one we will work with throughout our work, due to the two fundamental theorems of asset pricing.

Theorem 0.3 (First fundamental theorem of asset pricing)**.** *The market is arbitrage-free if and only if there exists a probability measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ such that the discounted price process of every tradable asset is a martingale with respect to $\mathbb{Q}$.*

Theorem 0.4 (Second fundamental theorem of asset pricing)**.** *The market is complete if and only if there exists a unique risk-neutral probability measure.*

With these premises, we then characterize our strategy as the minimization of the embedded risk in the following hedging portfolio:

$$P_{t,\tau} = V^{\Psi}_{t,\tau} + \pi^1_{t,\tau}\xi^1_{t,\tau} + ... + \pi^q_{t,\tau}\xi^q_{t,\tau} + \pi^B_{t,\tau}B_{t,\tau}, \qquad \pi^i_{t,\tau} = -\frac{\partial V^{\Psi}_{t,\tau}}{\partial \xi^i_{t,\tau}},\ i = 1, ..., q \tag{16}$$

The value $V^{\Psi}_{t,\tau}$ and $B_{t,\tau}$ corresponding to the option with payoff ϕ and the bond prices. To compute the Greeks $\pi^i_{t,\tau}$ we resort to our mathematical results and rearrange the risk-neutral density as a weighted linear combination of its moments $M^k_{t,\tau}$

$$f^{\mathbb{Q}}_{t,\tau}(x) \approx f_n(x) = \phi(x) \sum_{k=0}^{n} M^k_{t,\tau} \left(\sum_{i=k}^{n} \sum_{j=i}^{n} w_{i,k} w_{j,k} x^j \right) \tag{17}$$

From this relation the option value becomes

$$V^{\Psi}_{t,\tau} \approx H^{\Psi}_0 + M^1_{t,\tau}H^{\Psi}_1 + ... + M^n_{t,\tau}H^{\Psi}_n, \quad H^{\Psi}_k = \sum_{i=k}^{n} \sum_{j=i}^{n} w_{i,k} w_{j,k} \int_0^{+\infty} x^j \Psi(x)\phi(x)dx \tag{18}$$

This means that for a given set of strikes $K_1, ..., K_p$ we can then consider the vector of respective call options $C_{t,\tau} = [C^{K_1}_{t,\tau}, ..., C^{K_p}_{t,\tau}]'$ and the $p \times n+1$ matrix H with components $H_{i,j} = H^{K_i}_j$, to write the following approximations

$$C_{t,\tau} \approx HM_{t,\tau}, \qquad \frac{\partial C_{t,\tau}}{\partial \xi_{t,\tau}} \approx H \frac{\partial M_{t,\tau}}{\partial \xi_{t,\tau}} \tag{19}$$

This means that to estimate the Greeks we only need to find a way of expressing the partial derivatives of the risk-neutral moments. To do this we consider the following contracts.

Definition 0.5. Given an asset Z_t, the derivative contracts with maturity $t+\tau$ and values

$$F_{t,\tau} = \mathbb{E}^{\mathbb{Q}}[\Psi_1(Z_{t+\tau})] = \mathbb{E}^{\mathbb{Q}}[Z_{t+\tau}|\mathcal{F}_t]] \tag{20}$$

$$VS_{t,\tau} = \mathbb{E}^{\mathbb{Q}}[\Psi_2(Z_{t+\tau})] = -\frac{2}{\tau}\mathbb{E}^{\mathbb{Q}}[\log\left(\frac{Z_{t+\tau}}{F_{t,\tau}}\right)|\mathcal{F}_t] \tag{21}$$

are respectively called its *future* and its *variance swap*. Their Greeks are defined as

$$\Delta^K_{t,\tau} := \frac{\partial C^K_{t,\tau}}{\partial F_{t,\tau}} \qquad \nu^K_{t,\tau} := \frac{\partial C^K_{t,\tau}}{\partial VS_{t,\tau}} \tag{22}$$

and they are respectively called *delta* and *variance swap vega.*

We then can employ these derivative securities for this key relation

Proposition 0.6. *Given an asset Z_t, let $M^k_{t,\tau}$ be its k-th risk-neutral moment, $F_{t,\tau}$ its future and $VS_{t,\tau}$ its variance swap. Given the coefficients*

$$\beta_k = \frac{1}{VS_{t,\tau}} \log\left(\frac{M^k_{t,\tau}}{F^k_{t,\tau}}\right), \quad \forall k \in \mathbb{N} \tag{23}$$

We have that the following approximation holds :

$$M^k_{t,\tau} \approx F^k_{t,\tau} e^{\beta_k VS_{t,\tau}}, \quad \forall k \in \mathbb{N} \tag{24}$$

From this relation we can find the explicit definition of the moments derivatives leading to the following Greeks expressions

$$\Delta_{t,\tau} \approx HD, \quad \nu_{t,\tau} \approx HW \tag{25}$$

Where

$$D[k] = \frac{\partial F^k_{t,\tau} e^{\beta_k VS_{t,\tau}}}{\partial F_{t,\tau}} = kF^{k-1}_{t,\tau} e^{\beta_k VS_{t,\tau}}, \tag{26}$$

$$W[k] = \frac{\partial F^k_{t,\tau} e^{\beta_k VS_{t,\tau}}}{\partial VS_{t,\tau}} = \beta_k e^{\beta_k VS_{t,\tau}} \tag{27}$$

To compute the β_k we need a direct estimation of the moments which is achieved by performing a principal component regression and compute the coefficients

$$[\hat{c}^1_{t,\tau}, ..., \hat{c}^n_{t,\tau}] = \arg\min_{c_1,...,c_n} (\mathbf{Y}^* - \mathbf{X}c)'(\mathbf{Y}^* - \mathbf{X}c) \tag{28}$$

Where $\mathbf{Y}^* = \mathbf{Y} - \mathbf{X}_0$ and

$$\mathbf{Y} = [C^{K_1}, ..., C^{K_p}, P^{K_1}, ..., P^{K_p}]' \tag{29}$$

$$\mathbf{X}_0 = [A_0^{K_1}, ..., A_0^{K_M}, B_0^{K_1}, ..., B_0^{K_M}]' \tag{30}$$

$$\mathbf{X} = \begin{bmatrix} A_1^{K_1} & ... & A_n^{K_1} \\ . & . & . \\ . & . & . \\ A_1^{K_M} & ... & A_n^{K_M} \\ B_1^{K_1} & ... & B_n^{K_1} \\ . & . & . \\ . & . & . \\ B_1^{K_M} & ... & B_n^{K_M} \end{bmatrix} \tag{31}$$

Given the polynomial coefficients matrix

$$\mathbf{T} = \begin{bmatrix} w_{1,1} & 0 & ... & 0 \\ w_{2,1} & w_{2,2} & ... & 0 \\ . & . & . & 0 \\ w_{n,1} & w_{n,2} & ... & w_{n,n} \end{bmatrix} \tag{32}$$

The risk-neutral moments can then be derived by exploiting the relation

$$\hat{c} = \mathbf{T} M_{t,\tau} \Rightarrow M_{t,\tau} = \mathbf{T}^{-1} \hat{c} \tag{33}$$

Having developed our methodology for the computation of the Greeks we move to the last part where the empirical results are being reported and studied. We start with a brief retrospection over the SPX and the VIX dissecting their expressions to cast a light on the rationale of the measures. In particular, to understand the VIX form we show how it estimates the cumulative variance in relation to the stochastic model

$$d \ln S_t = \left[r - q - \frac{V_t}{2} + \lambda \left(\mu_J - e^{\mu_J + \frac{\sigma_J^2}{2}} \right) \right] dt + \sqrt{V_t} dW_t + J_t dN_t - \lambda \mu_J dt \tag{34}$$

$$dV_t = -\Delta_v \theta dt + \upsilon V_t^{\gamma} dB_t^* \tag{35}$$

Furthermore we prove the key equation

$$\mathbf{VIX}_t^2 10^4 \approx VS_{t,\tau} \tag{36}$$

confirming the functional link between the variance swap and the second risk factor.

We then move on to support the primary hypothesis that the major determinants for the behaviour of the option should be addressed to the futures and the variance swap on the underlying. To do this we perform a principal component analysis (PCA) over a panel of 1-month maturity SPX call options taken from the OptionMetrics platform, categorized in a pool of 9 moneyness

levels and collected at monthly frequency from 2006 to the first 4 months of 2016. We then prove that the leading drivers for SPX options are 2 and that they are functionally tied to the aforementioned derivative contracts. This is indeed clear after observing the very high correlations that the principal components have with respectively the SPX and the VIX^2 time series.
After proving also the optimal of degree of efficiency of our functional approximation linking the moment to the contracts values, we enter in the main section, where a consistent set of results concerning the risk-neutral density shape in different circumstances is provided together with a guide to interpret the market according to its properties. This includes the following

1. *Mean and standard deviation.* The first two risk-neutral moments carry information on the market expectation on the future underlying value and on the degree of related certainty. Since the terms are always conditional on the present available knowledge, nested in the filtration $\mathcal{F}_t$, we usually observe that the RND volatility depends on the maturity of the options it is extracted from. The closer the traders are to the expiration date the more certain they are on the future outcome, the narrower the density. This is also the reason why we denote it as $f^{\mathbb{Q}}_{t,\tau}$, including the horizon τ in the definition. This additionally suggests that the more ebullient the times and thus the larger the uncertainty, the more platykurtic the distribution.

2. *Asymmetry measures : skewness, (mean-mode)/standard deviation, (mean-median)/standard deviation.* Asymmetry in the distribution ascribes to the perception that the underlying has more chance to exceed (or fall behind) the most anticipated value, than instead falling behind it (or exceeding it). Furthermore a rapid truncation in one of the tail could mean that the market has set a limit in the believed decline or rise of the asset. Besides, as already mentioned, skewness could be regarded as a measure of the detachment from the Black-Scholes assumptions of log-normality for the returns and Brownian motion for the underlying stochastic process. To track dynamically the symmetry fluctuations, a good procedure could be the study of the pattern of the quantiles and their correlation with the market indices.

3. *Kurtosis.* The standardized fourth risk-neutral moment reflects the fatness of the tails, thus revealing the risk of extreme movements in the underlying, an indicator that could be used to evaluate the market anticipation of a financial shock.

4. *Modes.* The modes of a continuous distribution are its local maxima i.e. the set of most likely values. A multimodal density describes an heterogeneity of beliefs and thus a scattered probability, where the expectations of the agents are highly dispersed around the main value. On the contrary, a unimodal density could indicate a homogeneous convergence of the traders on the shape of the underlying future distribution.

In the last section we illustrate how the hedging performance could be measured in a sufficiently extended time span and present some remarks on how to possibly solve the technical issues emerging in low volatility regimes. Given a set of strikes $K^1, ..., K^M$ and time values $t_1 = t_0 + \frac{\tau}{N}$,...,$t_N = t_0 + \tau$, assuming a periodic rebalancing we define the replication squared error (RSE) and the benchmark replication squared error (BRSE), measuring the hedging performance of the practitioners' Black-Scholes strategy (PBS), as

$$RSE_i^{K_j} = \left(\Delta_{i-1}^{K_j}(F_i - F_{i-1}) + \nu_{i-1}^{K_j}(VS_i - VS_{i-1}) - (C_i^{K_j} - C_{i-1}^{K_j})\right)^2 \tag{37}$$

$$BRSE_i^{K_j} = \left(\Delta_{i-1}^{BS,K_j}(F_i - F_{i-1}) - (C_i^{K_j} - C_{i-1}^{K_j})\right)^2 \tag{38}$$

Where Δ_i^{BS,K_j} is the Black-Scholes delta corresponding to the value

$$\Delta_i^{BS,K_j} = \frac{\partial C_i^{K_j}}{\partial F_i} = \phi\left(\frac{\ln(\frac{F_i}{K_j}) + (r + \frac{\sigma^2}{2})(T - t_i)}{\sigma\sqrt{T - t_i}}\right) \tag{39}$$

with r risk-free interest rate and ϕ is the standard normal cumulative distribution function. From the above terms we can then define the hedging gain

$$G = \frac{1}{M}\sum_{j=1}^{M}\left(1 - \frac{\sum_{i=1}^{N} RSE_i^{K_j}}{\sum_{i=1}^{N} BRSE_i^{K_j}}\right) \tag{40}$$

The better the hedging strategy with respect to the benchmark Black-Scholes hedging, the closer to 1 the gain.
The empirically observed rising correlation between the SPX and the VIX,

emerging in low volatility regimes, could severely negatively affect the hedging. Since, indeed, we minimize the risk by taking direct positions on both the driving factors, we should always keep an eye on their covariance which, for reasons related to overly favorable monetary policies or market conditions, could significantly soar compromising the necessary condition for the mixed partial derivatives to be null and the hedging to work. This could be viewed as the VIX no longer translating the actual risk intrinsic in the products, an occurrence which causes a distorted perception of the market movements and induces the agents to over-leverage themselves towards a sudden unpredicted crash. We should therefore expect our hedging to be optimal in the highest turmoil, where the VIX adequately responds to the havoc, and badly in the quiet times, with the PBS turning out to be more solid as the options risk is loaded more on the delta. The methods propose to overcome this complication consist either in orthogonalizing the VIX with respect to the SPX, thus working the Greek of its perpendicular component treated as the risk factor, or using ATM options to indirectly hedge on the vega, exploiting the fact that these kind of options exhibit very little sensitivity to changes in the futures. Due to potential troubles arising in the first strategy the second emerges as more promising and easier to apply.