DOCUMENT DE SYNTHÈSE POUR
L'HABILITATION À DIRIGER LES RECHERCHES

Spécialité: Gestion

présentée par Guillaume Coqueret

au Laboratoire COACTIS de l'Université Lyon 2

Quelques contributions en finance quantitative

Soutenue le 10-01 2021 devant le jury suivant:

Catherine Bruneau	Rapporteure	Université Paris I Panthéon Sorbonne, Paris, France
Jean-Paul Laurent	Examinateur	Sorbonne School of Management, Paris, France
Olivier Le Courtois	Garant	EMLYON Business School, Lyon, France
Franck Moraux	Rapporteur	Université Rennes 1, Rennes, France
Jean-Luc Prigent	Rapporteur	Université de Cergy, Cery, France
Patrick Roger	Examinateur	EM Strasbourg, Strasbourg, France

Remerciements

La rédaction du présent document et la soutenance qui en est l'aboutissement doivent énormément à mon collègue et garant Olivier Le Courtois. Je le remercie pour le temps et l'attention qu'il a accordé à ce projet.

Contents

Introduction

Introduction générale

Le but de cette introduction (la seule section de ce mémoire rédigée en français) est de présenter le cheminement qui a amené à la rédaction de ce document. Ce parcours est dû autant à des préférences personnelles, qu'à des rencontres, qu'à des choix professionnels marquants.

Au début des années 2000, en école de commerce, un seul choix s'impose lorsque l'on souhaite combiner la gestion avec les disciplines quantitatives: la finance.[1] Cela étant dit, la finance quantitative est globalement et paradoxalement peu enseignée en profondeur en école de commerce et les formations d'excellence se trouvent principalement à l'université (Paris VI, Paris VII, Dauphine, Marne-la-Vallée et Évy, pour ne citer que quelques exemples franciliens). Une des raisons est que la filière quantitative s'appuie beaucoup sur le calcul stochastique, lequel n'est que marginalement enseigné en école de commerce.

Suivant l'ex DEA (Master 2 Recherche) de Paris VI (dit "El Karoui"), et sous l'influence de mon ancien professeur de probabilités (Thomas Simon, de l'université de Lille), je me passionne rapidement pour les processus de Lévy, sur lesquels j'ai fait un mémoire lié au pricing d'options simples lors d'un précédent Master de Recherche à l'Université Paris I. Ma thèse, encadrée par Patrice Poncet (ESSEC) et Thomas Simon (Université Lille-1) est focalisée sur ce sujet; le but étant de développer mes compétences techniques, lesquelles me seront utiles, que ce soit dans le monde académique ou dans les salles de marché.

La tâche est cependant ardue car les modèles dits "exponentiels de Lévy" sont trés populaires à ce moment (circa 2008), et il n'est pas aisé de trouver des spécifications qui soient à la fois inexplorées et accessibles analytiquement. Ma thèse s'articule donc autour de cette nécessité de trouver des processus suffisamment compliqués pour qu'ils soient utiles et suffisamment simples pour qu'on puisse dériver des formules élégantes. Les problèmes les plus difficiles sont ceux liés au pricing d'options dites "exotiques", c'est à dire, dépendant des trajectoires des sous-jacents. Ceux-ci peuvent être résolus avec certaines identités particulières lorsque le payoff dépend des extremas du sous-jacent. Dans le cas des options de type barrière, cela réduit le

[1] Ce n'est plus vrai en 2021: les data sciences fournissent des débouchés très intéressants pour les filières commerciales et concurrencent désormais largement la finance comme choix de sortie (et d'orientation) pour nombre d'élèves.

prix à payer par l'acquéreur, ce qui rend ces produits plus attractifs, bien que plus complexes.

C'est cette voie de recherche que j'emprunte alors, à travers le prisme des méthodes dites de "Wiener-Hopf". Un avantage est que dans un cadre probabiliste les formules analytiques peuvent être vérifiées via des méthodes numériques, lesquelles sont indispensables en pratique et utilisées par tous les vendeurs de produits dérivés. Les simulations de Monte-Carlo, quoique souvent lentes, sont très flexibles et permettent d'approximer les prix de tous ces produits. Cette technique est donc un outil très utile pour valider tous les résultats théoriques et je l'ai utilisée fréquemment lors de mes travaux doctoraux.

Ma thèse et mes trois premiers articles publiés s'articulent autour des notions très proches de fonction caractéristiques et de transformées de Laplace. Deux d'entre eux traitent de formules fermées (plus élégantes), et le premier mentionne une technique d'approximation d'inversion de la transformation de Laplace. Cette dernière est souvent utile pour obtenir le prix d'options à maturité fixe, lesquelles sont les seules vendues sur les marchés.

À la toute fin de ma thèse, j'accepte une proposition pour un poste d'ingénieur quantitatif dans le centre de recherche de l'EDHEC-Risk Institute (ERI). La mission principale de ce poste est de participer à la réalisation de chaires de recherche financées par les partenaires de l'institution - et d'écrire des articles académiques.

Les problématiques du centre sont alors très différentes et ma recherche va prendre un virage prononcé. Les sujets de pricing, liés à des activités dites de "sell-side" (les opeérateurs qui vendent des dérivées) font place à des besoins liés au "buy-side" (les gérants d'actifs au sens large), c'est à dire des problèmes de choix de portefeuille.

Les outils mathématiques sous-jacents des deux activités sont complèment différents:

- pour le **pricing**, l'accent est mis sur le calcul stochastique, l'analyse et le contrôle optimal, les méthodes numériques de simulations et d'équations aux dérivés partielles (Monte-Carlo et différences finies notamment);
- pour l'**allocation d'actifs**, on a besoin d'optimization, d'algèbre (calcul matriciel), de statistiques, d'asset pricing (en temps discret ou continu),[2] et de techniques de backtest.

Intellectuellement, le changement de paradigme fut très stimulant, d'autant plus que je me rendis compte que je préferais finalement les outils liés à mes nouvelles attributions. J'ai travaillé principalement sur 3 projets lors de mes 27 mois à l'ERI:

1. L'allocation d'actifs, notamment à travers la notion de Factor Risk Parity, qui est telle que la contribution factorielle sous-jacente soit égale entre tous les facteurs de risque.
2. La production de portefeuilles d'actions qui soient le plus proche en terme de risque et de performance des portefeuilles d'obligations (une idée surprenante mais assez fertile et très utile pour certains gérants d'actifs).

[2] Formellement, il est vrai que le pricing de produits dérivés est aussi une sous-branche de l'asset pricing.

3. Le factor timing, c'est à dire, l'allocation dynamique à des portefeuilles factoriels en fonction de variables macro-économiques.

Ce passage a donc réorienté mes intérêts en recherche dans une direction résolument axée vers les problèmes d'allocation, de statistiques, et d'optimisation.

Mon temps à l'ERI fut relativement bref car j'aspirais à plus de liberté dans ma recherche, et je fus heureux d'accepter un poste d'assistant professeur de finance à la Montpellier Business School. C'est à cette période que je décidais d'entamer un nouveau cycle de recherche qui allait m'occuper sans discontinuer depuis. Une synthèse de mes cycles est fournie dans la Figure 0.0.1.

Mon passage à l'EDHEC me fit réaliser à quel point les modèles en finance étaient gourmands en données et c'est pourquoi la science, relativement jeune, qui leur est dédiée m'apparu un choix évident. Je changeais de langage de programmation de Matlab à R et je commençais à me former aux techniques d'apprentissage automatique (machine learning - abrégé ML dans le reste du présent document).

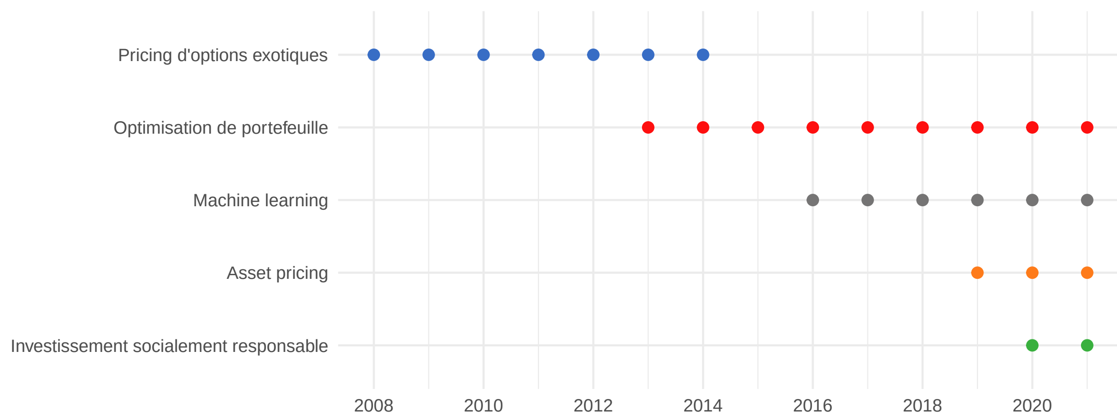

Figure 0.0.1: **Sujets de recherche par année**.

Cette discipline, je le compris vite, n'était pas totalement incompatible avec la gestion de portefeuille et la combinaison des deux est un sujet sur lequel j'ai écrit un livre et qui pose plus de questions qu'il n'en résoud. L'application la plus directe du ML en théorie du portefeuille et en asset pricing est la prédiction des rendements des actifs pour lesquels on est capable d'amasser de larges quantités de données (on fait souvent réfréence au terme "*Big Data*"). Ces dernières sont utilisées comme matière première pour des modèles prédictifs par des chercheurs et analystes, lesquels, tentent d'en extraire le sens des marchés. L'exercice est cependant rendu compliqué par les relations toujours changeantes entre les variables dépendantes et indépendantes des modèles utilisés.

Les intéractions entre l'apprentissage supervisé, le big data et l'asset pricing composent la majeure partie de ma recherche depuis mon arrivée à l'EMLYON. La capacité des modèles à **généraliser**, c'est à dire à prédire l'avenir grâce aux observations passées demeure pour moi un mystère et un sujet d'étude passionnant (que je détaille à la fin de ce document). Fondamentalement, cela nécessite de déterminer des relations de causalité, lesquelles sont notoirement difficiles à prouver dans leur sens le plus strict (par exemple, celui de Pearl).

Progressivement, je m'éloigne de la technique statistique pure pour me recentrer sur la compréhension des marchés et la création de modèles d'asset pricing en temps discret plus conventionnels mais plus élégants, surtout lorsqu'ils émanent d'équilibres économiques. Je travaille actuellement sur ce type de modèles, où des agents ont des demandes inhomogènes concernant les caractéristiques des actifs, ce dont on infère des liens entre rendements et ces mêmes caractéristiques.

Enfin, en 2020, ayant été sollicité par une institution financière pour développer une offre de fond d'investissement "vert", lesquels ont connu une demande très soutenue ces dernières annés, je me suis mis récemment à approfondir mes connaissances en **finance durable** et responsable. Le sujet est tentaculaire et chacune de ses facettes comporte de multiples ramifications.

Ayant voulu commencer par la construction de portefeuilles verts, j'ai réorienté ma recherche pour analyser le lien entre le classement environemental, social et de gouvernance (ESG) des sociétés d'un côté et leur performance financière de l'autre. Puis, j'ai examiné les sujets de risque climatique, les normes de reporting (ou leur absence), les modèles théoriques d'évaluation économique du coût des émissions de carbone, etc. Il est de facto inutile de vouloir ne comprendre qu'un périmètre restreint du problème. Cependant, le bénéfice d'une compréhension large est chronophage.

Le sentiment d'urgence qui s'en dégage rend l'exercice toutefois plus qu'intéressant, d'autant plus que la gravité de la situation climatique est relayée de maniére souvent peu scientifique dans les médias. Il est appréciable d'avoir quelques notions pour lire les annonces et reportages avec un minimum de recul. Très récemment, ayant été contacté par une entreprise spécialisée du secteur, je vais entamer une étude sur l'importance du Scope 3 des émissions carbonées pour la construction de portefeuille.

Asset pricing théorique et finance durable sont les deux thèmes qui devraient m'occuper pour les quelques années à venir. En parallèle de ces nouvelles directions de recherche, j'ai été plusieurs fois sollicité ces dernières années pour encadrer des thèses. J'ai donc senti qu'il était pour moi venu le temps de m'investir pour rédiger mon habilitation.

Échantillon de références académiques

Les contributions sont listés en ordre chronologique inversé et groupées par thème. Elles figurent également dans la bibliographie, ainsi que d'autres publications plus mineures et des articles non-encore acceptés.

Produits dérivés

- A note on implied correlation for bivariate contracts, **Economics Bulletin**, Vol. 40, No. 2 (2020), pp. 1388-1396 (with B. Tavin)
- An investigation of model risk in a market with jumps and stochastic volatility, **European Journal of Operational Research**, Vol. 253 (2016), pp. 648-658 (with B. Tavin)
- On the distribution of the supremum of the spectrally negative stable process with drift, **Statistics & Probability Letters**, Vol. 107 (2015), pp. 333-340
- Lookback option prices under a spectrally negative tempered-stable model, **International Journal of Theoretical and Applied Finance**, Vol. 16, No. 3 (2013)

Théorie du portefeuille

- Equity portfolios with improved liability-hedging benefits, **Journal of Portfolio Management**, Vol. 43, No. 2 (Winter 2017), pp. 37-49 (with L. Martellini and V. Milhau)
- Characteristics-based portfolio choice with leverage constraints, **Journal of Banking and Finance**, Vol. 70 (2016), pp. 23-37 (with M. Ammann and J-P. Schade)
- Diversified minimum-variance portfolios, **Annals of Finance**, Vol. 11, No. 2 (2015), pp. 221-241

Finance et science des données

- Machine learning for factor investing, **CRC Press**, 2020 (with T. Guida)
- Training trees on tails with applications to portfolio choice, **Annals of Operations Research**, Vol. 288 (2020), pp. 181-221 (with T. Guida)
- Stock-specific sentiment and return predictability, **Quantitative Finance**, Vol. 20, No. 9 (2020), pp. 1531-1551
- Approximate NORTA simulations for virtual sample generation, **Expert Systems with Applications**, Vol. 73 (2017), pp. 69-81

Modèles multi-agents

- Procedural rationality, asset heterogeneity and market selection, **Journal of Mathematical Economics**, Vol. 82 (2019), pp. 125-149
- Empirical properties of a heterogeneous agent model in large dimensions, **Journal of Economic Dynamics and Control**, Vol. 77 (2017), pp. 180-221

Chapter 1

Early work: pricing derivatives

1.1 Introduction

The topic of pricing and hedging derivative products has gained a lot of traction in the 1990 and 2000 decades, even though the early breakthroughs in the fields dated back to 1973 at least (Black and Scholes (1973) and Merton (1973) - henceforth, BSM).[1] In the sequel, we enumerate some reasons for these developments.

The valuation of a derivative is based on an underlying process which is usually modelled as S_t, where t is the index of time. The underlying is most often a unique financial reference (e.g., a stock price, a currency quote, etc.), but it can also be generalized to higher dimensions (e.g., for basket options which depend on two or more underlyings). The price of the product is equal to the discounted value of its payoff at maturity:

$$p(S, f, T, \mathbb{Q}) = \mathbb{E}_{\mathbb{Q}}\left[e^{-rT} f(S_s, 0 \le s \le T)\right], \tag{1.1}$$

where f is the payoff function which is possibly defined on the whole trajectory of the process and $\mathbb{Q}$ is a suitable probability measure. T measures the maturity of the derivative, that is, the horizon at which the final payoff is determined. For simplicity, we consider that the present time is fixed at $t = 0$ and that the discounting (risk-free) rate is constant and equal to r. Thus, three elements are required to compute the above integral:

1. The definition of the process S_t, e.g., via its probabilistic properties under $\mathbb{Q}$;
2. The payoff function f;
3. A numerical method to assess the integral - in the case where no analytical solution is available.

All of these elements have benefited from favorable trends in the 1990s-2000s:

1. Researchers have started to take interest in non-stationary processes with non-standard volatilities (Heston (1993), Dupire (1994)) and also in non-Brownian

[1] Not to mention the pioneering work of Bachelier (1900).

(Gaussian) assumptions. For the latter, we refer to the very complete monographs on applications of Lévy process for derivatives Schoutens (2003), Cont and Tankov (2003), Kyprianou et al. (2006) and, more recently, Eberlein and Kallsen (2019). This has opened the field to more realistic ways to model trajectories in continuous time. Lévy processes are very convenient because they allow for very flexible parametric forms and can be easily simulated (see Asmussen and Rosiński (2001), Kuznetsov et al. (2011)).

More complex diffusions (e.g., semimartingales, which are treated in detail in Jacod and Shiryaev (2013)) much harder to handle in all generality and have rarely been applied in practice. Indeed, parametric models (e.g., Heston (1993)) are mostly favored, even though the local volatility model of Dupire (1994), in which the volatility of the underlying depends on its level, may be considered an exception. All modern models are meant to generate realistic implied volatility surfaces.[2]

2. In the meantime, the appetite in the industry for more complex (customer-centric) products has also thrived. Exotic options (Asian, barrier, basket, Bermuda, binary, chooser, lookback, to name a few in alphabetical order) started selling on trading floors, and more and more engineers (so-called "*quants*") have been trained to handle (price and hedge) these complex instruments. The set of functions f in Equation (1.1) is now quite rich. Consequently, from an academic perspective, finding closed-form expressions for Equation (1.1) was (and remains!) an appealing challenge.
3. When analytical formulae are not available, the integral has to be computed numerically. The most flexible framework nests all tools related to Monte-Carlo simulations. However, for exotic (path-dependent) options, they are notoriously slow, because the convergence to continuous prices requires a very fine simulation grid *for each trajectory*.[3] Therefore, other techniques have been developed, notably relying on:
   - **partial differential equations**: Matache et al. (2004), Cont and Voltchkova (2005) and Achdou and Pironneau (2005);
   - **Malliavin calculus**: Fournié et al. (1999), Di Nunno et al. (2009);
   - **series representations** and **polynomial expansions**: Ackerer and Filipović (2020);
   - **Fast-Fourier transforms**: Carr and Madan (1999)), and Fang and Oosterlee (2009);
   - **Complex integration via residues**: Aguilar and Kirkby (2021)

[2] The seminal Black-Scholes-Merton (BSM) formula yields imperfect prices. Notably, it implies a constant volatility. In practice, to obtain realistic prices, the volatility has to be adjusted along the strike and maturity dimensions. The volatility value in the BSM model that is able to yield the exact market quote is referred to as the *implied volatility*..

[3] This can be partly alleviated via multilevel Monte Carlo, see Giles (2008), Ferreiro-Castilla et al. (2014), Giles and Xia (2017), to cite but a few.

- **Wiener-Hopf factorizations**: see the overview Kudryavtsev and Luzhetskaya (2020).

The combination of these three factors posterior to the 1990s decade has allowed the field to blossom and the dedicated literature has thrived since then. In this chapter, we summarize results in different directions. The first direction, in Section 1.2, pertains to the first and second point (exotic options with Lévy-driven underlying). When processes and payoffs are not standard, analytical formulae are much harder to attain. We provide developments for particular option types (barriers and lookbacks) when the underlying follows special Lévy processes. The second part (Section 1.3) relates to model risk. Because there is no one ultimate way to choose a process (or model), resulting prices are subject to risk. By definition, all models are wrong, but it is important (and interesting) to be able to quantify *how* wrong they can be. This is useful because it gives traders and quants the margin of error when pricing structured products. Finally, the last point we shed light on is the notion of *implied correlation* in Section 1.4. While implied *volatility* is a well-defined and well understood conception, implied correlation remains a more esoteric notion and we propose a rigorous way to define and evaluate it.

1.2 Some exotic options under particular Lévy models

The barrier and lookback options are two popular families of exotic derivatives that depend on the realized path of their underlying. In all generality, we assume that the latter can be modelled as an **exponential Lévy process**, $S_t = S_0 e^{X_t}$, such that its discounted value is a martingale under the reference probability. Barrier options allow to reduce the price of the derivative because payment of the payoff is conditional on some event involving the running supremum or infimum of the underlying. Thus, the payoff is paid in fewer realizations of the world, compared to a vanilla (non path-dependent) option. The lookback option is more expensive because it is designed to deliver a positive payoff (maximum minus terminal value or terminal value minus minimum).

The technical hurdle in the analytical computation of these options is twofold:

1. first, in the case of the lookback option, the distribution of the extrema are not known in all generality. If fact, in many cases, the distribution of the underlying can be also complex and only obtained via series expansions. The favorable feature is that it is not necessary to know the density of the random variables, because there Laplace transform (or, almost equivalently, their Fourier transform) is sufficient.

2. For barrier options, there is nonetheless one additional obstacle arises. Indeed, barrier options require the knowledge of the **couple** (underlying, extremum). The latter is very hard to characterize for finite times (one exception being Chaumont et al. (2013), but it is impossible to exploit numerically). The only information about the underlying-extremum pair that is "*easy*" to compute

is the so-called **Wiener-Hopf** factorization, which yield the bivariate Fourier transform (characteristic function) of the pair.[4] However, it only holds when the processes are stopped at an independent time which is exponentially distributed (i.e., this gives the Laplace transform in the time domain). Thus, in order to retrieve the desired density, a double inversion (Fourier and Laplace) must be performed, which is impractical. This is why either tricks are required, or closed-form results can only be obtained in very particular cases for which parametric identities can be derived.

Below, we consider two cases, which are detailed in the two subsections below. Both models originate from the finite log-stable process studied in Carr and Wu (2003). In the first part, we show how to compute the distribution function of the running supremum of the asymmetric stable process *with drift*. The drift part is critical because under the martingale measure, the process must be corrected by a negative drift. Analytically, this substantially complexifies the required algebra. We are nonetheless able to provide a tractable formula for the Up and In Put (UIP barrier option). In the second subsection, we work with a model that includes a mitigating parameter which tames the behavior of the Lévy measure towards its tails. This yields the so-called *tempered* stable model, which is a rich class of processes, though we only cover a special case below.

1.2.1 Barrier options and drifted asymmetric stable processes

The starting point is the spectrally negative α-stable ($\alpha \in (1,2)$) process $X^{(\alpha)} = \{X_t^{(\alpha)}, t \geq 0\}$, which has three defining features. First, it is a Lévy process, that is, its increments are independent and stationary and it starts at zero almost surely. Second, it is self-similar, i.e. $\{X_{ct}^{(\alpha)},\ t \geq 0\}$ has the same distribution as $\{c^{1/\alpha} X_t^{(\alpha)},\ t \geq 0\}$, for any $c > 0$. Third, it jumps only downwards. It is completely characterized by its Laplace transform

$$\mathbb{E}\left[e^{\lambda X_t^{(\alpha)}}\right] = \mathbb{E}\left[e^{\lambda t^{1/\alpha} X_1^{(\alpha)}}\right] = e^{t\lambda^\alpha}, \quad \lambda \geq 0. \tag{1.2}$$

This incredibly simply form masks very complicated identities for the density of the law of the process (see, e.g., Zolotarev (1986)). Note that when $\alpha = 2$, we recover the standard Brownian motion. The asymmetric case with $\alpha \in (1,2)$ is the only one that can be used for option pricing because it is the only one, which, when exponentiated, will yield finite prices. This comes from the very light right tail (low probability of large positive jumps), combined to a heavy left tail (high probability of negative jumps, akin to crashes).

Here, we are interested in the drifted version of $X_t^{(\alpha)}$, $X_t^{(\alpha,\mu)} = X_t^{(\alpha)} + \mu t$, where $\mu \neq 0$ - the case $\mu = 0$ is treated in Michna (2013) and Michna et al. (2015). In particular, we will focus on its running supremum $M_t^{(\alpha,\mu)} = \sup\limits_{0 \leq s \leq t} X_s^{(\alpha,\mu)}$. A simple

[4]For a probabilistic treatment on the Wiener-Hopf factors, we refer to Kyprianou (2006), and for an analytical perspective, we point to Sato (1999).

representation in the case $\mu = 0$ is given in Michna (2013), and we aim to propose an alternative representation when $\mu \neq 0$. The opposite problem when the process is spectrally positive is tackled in Michna (2011).

The main result can be stated as follows. If $\alpha \in (1,2)$ is not a rational number, then for $x, t > 0$,

$$F(x) = P\left[M_t^{(\alpha,\mu)} \leq x\right] = \frac{\alpha}{\pi}\sum_{n=1}^{\infty}(-1)^{n-1}\frac{\Gamma(n/\alpha+1)}{n!\,n}\sin\left(\frac{\pi n}{\alpha}\right)\,{}_2F_1\left(1-n,-\frac{n}{\alpha};1-\frac{n}{\alpha};\frac{\mu t}{x}\right)\frac{x^n}{t^{n/\alpha}}, \quad (1.3)$$

where ${}_2F_1$ is the hypergeometric function[5]

Even though the formula may appear cumbersome, it is relatively straightforward to compute, especially with modern programming languages (Matlab, Mathematica, R, Python) and their advanced numerical libraries. In particular, it is possible to split the series into two parts, one finite and one infinite. The finite one is easy to evaluate and the infinite one can be precisely approximated thanks to the asymptotic expansions of the hypergeometric function.

Subsequently, in order to convert this result into an element of a pricing formula, we must recall some basic identities with respect to vanilla and exotic put options. The former have the following payoffs: $P(S_T, K) = \max(K - S_T, 0)$, where T is the maturity and K is called the strike of the option. The fair value for such a financial product is hence $p(K,T) = \mathbb{E}\left[\max(K - S_T, 0)\right]$. The expectations are associated with $\mathbb{Q}$, but we omit this reference for notational ease (we will only work under $\mathbb{Q}$). These put prices depend solely on the distribution of the underlying S at maturity. We are then interested in one particular option, namely the Up and In Put, which requires that the underlying reaches some value before maturity. Its payoff is the following:

$$UIP(S_T, M_T, K, B) = \max(K - S_T, 0)\mathbb{1}_{M_T \geq B},$$

where M is the running maximum of S and B is a barrier value such that $0 < B < K$. We use the notation $\mathbb{1}_{\{A\}}$ for the indicator function of the set A. The trick is then to resort to the replication reasoning of Carr and Bowie (1994), according to which it can be shown that the issuing price of the UIP option is equal to

$$uip(K,T,B) = (K - B)P[M_T > B] + c(K,T), \quad (1.4)$$

where $c(K,T) = \mathbb{E}\left[\max(S_T - K, 0)\right]$ is the price of a vanilla call. This last formula is key, as it links the price of the UIP to a simple expression that involves $P[M_T > B]$, which is equal to one minus the cumulative distribution function of M_T - which is precisely the expression we have in Equation (1.3).

In the paper Coqueret (2015b), numerical examples are provided. In particular, the convergence from very discrete sampling (e.g., when extrema are observed monthly) to continuous trajectories (which are purely theoretical) is shown.

[5] See for instance chapter II in Bateman (1953)):

$$ {}_2F_1(a,b;c;z) = \sum_{n=0}^{\infty}\frac{(a)_n(b)_n}{(c)_n}\frac{z^n}{n!}, \quad |z|<1, \quad \text{with} \quad (q)_n = \begin{cases} 1 & n=0 \\ q(q+1)\cdots(q+n-1) & n \geq 1 \end{cases}. $$

Unfortunately, it is not possible to cover the full range of barrier option combinations (call/put, up/down, in/out) with the above results. Analytical formulae in the general case are probably out of reach.

1.2.2 Lookback options and asymmetric tempered-stable processes

The second configuration for which we are able to obtain a closed form for prices pertains to tempered stable processes. The origins of this family of processes in option pricing go back at least to Carr et al. (2002) for vanilla products. The tempered stable distribution has other applications in finance (Bianchi et al. (2010), Kim et al. (2010), Fallahgoul and Loeper (2021)).

In this section, the model reads

$$S_t = S_0 e^{(r-d-(\sigma+\lambda)^\alpha+\lambda^\alpha)t+\sigma X_t}, \quad t \geq 0, \tag{1.5}$$

with $r-d = (\sigma+\lambda)^\alpha - \lambda^\alpha$ so that the martingale condition is satisfied. The lookback prices we are interested in are

$$LC = \begin{cases} e^{-rt} S_0(\mathbb{E}[e^{\sigma X_t}] - \mathbb{E}[e^{\sigma I_t}]), & \sigma > 0 \\ e^{-rt} S_0(\mathbb{E}[e^{\sigma X_t}] - \mathbb{E}[e^{\sigma M_t}]), & \sigma < 0 \end{cases}, \quad LP = \begin{cases} e^{-rt} S_0(\mathbb{E}[e^{\sigma M_t}] - \mathbb{E}[e^{\sigma X_t}]), & \sigma > 0 \\ e^{-rt} S_0(\mathbb{E}[e^{\sigma I_t}] - \mathbb{E}[e^{\sigma X_t}]), & \sigma < 0 \end{cases}$$

where

$$I_t = \inf_{0 \leq s \leq t} X_s, \quad M_t = \sup_{0 \leq s \leq t} X_s;$$

LP and LC are, respectively, the issuing prices of the continuously monitored lookback put and lookback call written on the stock S with maturity t. The cases $\sigma < 0$ we inferred from the identities $\sup\limits_{0 \leq s \leq t} - X_s = -I_t$ and $\inf\limits_{0 \leq s \leq t} - X_s = -M_t$. Both LP and LC depend on α, σ, λ and t, but for notational convenience, we will henceforth omit this dependency.

The expectations related to the underlying itself (X_t) are straightforward to compute: the hard part comes from the extrema. The key ingredient that unlocks the problem of the Laplace functions is the Wiener-Hopf formula which we can invert for asymmetric tempered-stable processes.[6] When inverting the Wiener-Hopf factors in the time-domain, we are able to show that the "*simple*" Laplace transforms of M_t and $-I_t$ are given by

$$\mathbb{E}[e^{-\sigma M_t}] = 1 - \sigma \int_0^t e^{-\lambda^\alpha x} x^{1/\alpha - 1} E_{\frac{1}{\alpha},\frac{1}{\alpha}}((\lambda - \sigma) x^{1/\alpha}) dx, \quad \sigma \in \mathbb{R}, \tag{1.6}$$

and for $\sigma + \lambda > 0$,

$$\mathbb{E}[e^{\sigma I_t}] = e^{t((\sigma+\lambda)^\alpha - \lambda^\alpha)} \left(1 - \sigma \int_0^t e^{-(\sigma+\lambda)^\alpha x} x^{1/\alpha - 1} E_{\frac{1}{\alpha},\frac{1}{\alpha}}(\lambda x^{1/\alpha}) dx\right), \tag{1.7}$$

[6] In similar contexts, the Wiener-Hopf factorization is also exploited in Kuznetsov et al. (2011), Hackmann and Kuznetsov (2014), Fusai et al. (2016)

where $E_{a,b}(\cdot)$ is the two parameter Mittag-Leffler function:

$$E_{\alpha,\beta}(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\beta + \alpha k)}, \quad \alpha > 0, \quad \beta, z \in \mathbb{R}.$$

One very favorable feature of these expressions is that they can be approximated by discretizing the integral and computing the tail via the asymptotic behavior of the Mittag-Leffler function. Also, our result makes it possible (though with some effort) to evaluate the convergence speed of discrete monitoring towards continuous monitoring. Indeed, we can prove that

$$LC - LC^{(n)} = O(n^{-1/\alpha}), \quad LP - LP^{(n)} = O(n^{-1/\alpha+\epsilon}), \epsilon > 0, \quad n \to \infty,$$

where $LC^{(n)}$ and $LP^{(n)}$ are the lookback option prices such that the extrema of the underlying are observed n times during the life of the option (the monitoring dates are assumed to be uniformly spaced). What this shows is that the $\alpha-$index of the process is key to the convergence. A small α (close to 1) warrants a faster rate. Using standard identities such as the put-call parity, we are then able to recover other prices, but, as is often the case, some configurations remain unknown in closed form.

1.3 Model risk

The previous section illustrated the importance of the choice of a parametric process that models the dynamic behavior of the underlying. There no unambiguous way to choose a "*best*" model a priori. All models end up being wrong, but what matters is the magnitude and frequency of errors. One issue is for instance to make sure that a model satisfies the empirical properties observed in practice in financial time series (see Cont (2001)). Another common practice is to make sure that the model is able to match prices observed on option markets. This requires a critical step, known as **calibration**, but even after an efficient calibration, risks remain (Guillaume and Schoutens (2012)).

There are several reasons for this. One the one hand, it is hard to fit the implied volatility on the whole surface, and there are often parts for extreme (high or low) values of the strike or the maturity that remain approximatively priced. On the other hand, a calibration on model prices will reflect the market expectations at a given time, but they can change if market conditions or underlying characteristics fluctuate.

We briefly describe the theoretical setup below. Let Z be a contingent claim with a single payoff paid at time T and denoted by Z_T. Z_T is a random variable on some probability space $(\Omega, \mathcal{F}_T, \mathbb{P})$. Z is said to be path-dependent if its payoff function, denoted by z, is a function of the trajectory of S, its underlying, on $[0, T]$, that is if $Z_T = z(\{S_t, t \in [0, T]\})$. At initial time, a no arbitrage price of Z is given by

$$Z_0 = \mathbb{E}^{\mathbb{Q}}\left[e^{-rT} z(\{S_t, t \in [0, T]\})\right]. \tag{1.8}$$

When the payoff is a function of the final asset price only, the contingent claim is said to be European and $Z_T = z(S_T)$. We assume that vanilla options (calls and puts) written on S are available for a finite number of strikes and maturities. $C(K,T)$ denotes the call option struck at K and with maturity T. Its payoff is written $C_T(K,T) = (S_T - K)^+ = \max(S_T - K, 0)$ and its initial price is $c(K,T)$ (following our notations from the previous section).

It is important to understand the implications of Equation (1.8). If a public price Z_0 is available for the payoff z, then this can be seen as valuable information about the underlying S. Indeed, if many calls and puts have quotes for maturity T then these prices can be used to determine the shape of the distribution of S_T, via Equation (1.8) (only a few distributions can correctly match market prices, see for instance Jackwerth and Rubinstein (1996)).

Now, even more interestingly, when working with path-dependent options, the information embedded in the prices can be used to characterize not only the distribution of S at a fixed time, but also *through* time. This is because the payoff depends on the whole trajectory of the underlying until maturity.

Lets us now present a very short exposé of model risk in the field of derivatives.[7] The agent (trader, quant, researcher, etc.) has at her disposal a dataset $\mathcal{D}$ which contains information about the underlying (prices, implied volatility surfaces,[8] exogenous derivatives prices, etc.). She wishes to calibrate a parametric model $\mathcal{M}$ using an estimation method $\mathcal{E}$. Thus, for a given product with specified payoff, the price will depend on these three elements, i.e., $Z = Z(\mathcal{D}, \mathcal{M}, \mathcal{E})$. Consequently, two agents who have different data, model preferences and/or estimation techniques will likely end up with two very different prices - because their estimated models will be different. This is what is commonly referred to as *model risk*.

In all generality, it is very hard to identify and quantify the relative importance of these sources of model risk. This is because the difference $Z(\mathcal{D}, \mathcal{M}, \mathcal{E}) - Z(\mathcal{D}', \mathcal{M}', \mathcal{E}')$ embeds mitigating effects across all components. Thus, it is easier to only make one component vary. In Coqueret and Tavin (2016) for instance, it is the parametric families that are considered, which means that the risk comes from $\mathcal{M} \neq \mathcal{M}'$. The error is quantified as the maximum pricing error that is obtained while spanning 21 parametric families of models. More recent contributions (e.g., Lazar et al. (2020)) are more ambitious and try to disentangle the effects of several sources of risk (e.g. both model specification risk and estimation risk).

Naturally, the impact of models, estimation techniques, etc., is also contingent on the specifications of the product. Depending on the nature of the product (e.g., variance swap versus forward-start option), the maturity may matter or not, the market situation (calm or turbulent), the market itself (US versus Europe), may also have an impact. For example, in Coqueret and Tavin (2016), we show that model risk in forward start options in overall increasing with the maturity of the

[7] This short summary, while originating from the paper Coqueret and Tavin (2016), is also inspired and influenced by more recent research, notably Lazar et al. (2020).

[8] Implied volatilities are often preferable because they are comparable in scale across assets and because they roughly keep the same magnitude over the whole surface. This is typically not true for prices, which may generate problems in the calibration phase.

option.

While unconditional results cannot be obtained, the important message is that, in practice, agents operating with such derivatives and models should be aware that their modelling choices are imperfect and can have an impact on the P&L of their trading portfolios. One possible way to mitigate the model risk would be to resort to ensembles (averages of various models), but this direction has not yet been much investigated in the literature.

1.4 A formal definition of implied correlation

As mentioned in the sections above, the implied volatility is a tool that is not only mainstream, but also ubiquitous in the derivatives industry (and in the articles of academics who work on these topics). Simply put, the implied volatility is the value of the volatility parameter which, if inserted in the Black-Scholes formula, yields the price of the vanilla option (put or call) observed on the market (all other parameters being naturally observable).

Naively, it seems tempting to apply the definition of implied volatility to the notion of implied correlation. This is however less straightforward because it of course requires two underlying assets (or more). The technical hurdle is that even if a pair of assets is assumed to have Gaussian marginals, the correlation structure remains an open choice. One natural choice (see, e.g., Coqueret and Tavin (2020)) is to push the Gaussian assumption one step further and to assume that the dependence relationship is driven by a Gaussian copula.

The next step is to identify derivative payoffs that take into account two series as underlying. At least four families can enter this categoy: basket options, spread options, maximum options and double binary options (we refer to Coqueret and Tavin (2020) for their relative payoffs). All of them fall under the umbrella of the so-called $\mathcal{I}$ (integral) class of options, which have a particular kind of integrated functional form for their payoffs. This family encompasses the two following forms for payoffs z depending on two variables: $z(s_1, s_2) = \mathbf{1}_{\{\pm s_1 \le K_1, \pm s_2 \le K_2\}}$ and, for two well-defined continuous functions g_1 and g_2, $z(s_1, s_2) = \int_L^U \mathbf{1}_{\{\pm s_1 \le g_1(x), \pm s_2 \le g_2(x)\}} dx$.

It is then possible to show that

- for this family of payoffs,
- under an assumption of Gaussian copula structure,
- and when the marginal distributions of the underlying are known,

then there is a unique correlation value for which the theoretical model can match the observed market price. This value can be called **implied correlation**. The above assumptions are fairly loose, which implies that the results hold in relatively general contexts.

It is important to underline that our definition of implied correlation relies on the correlation that is priced in a multivariate contract. Other contributions, such as Buss and Vilkov (2012), work with *simple* (i.e., univariate) option prices, which only capture the dynamic of the underlying. Their definition of implied correlation comes from the cross-section of options and the related dependence structures (see Figure 1.4.1).

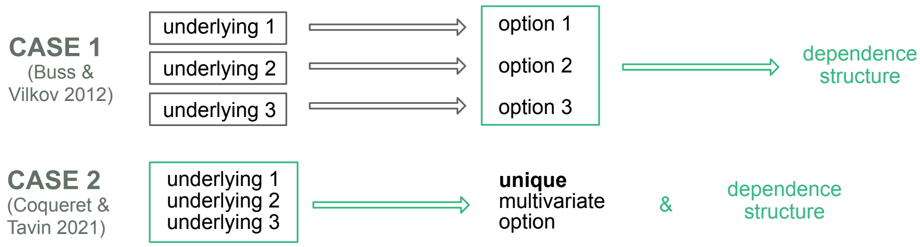

Figure 1.4.1: **Two views on inference on the dependence structure**.

Simply put, the correlation then comes from the correlations implied from option values across assets, whereas in our case the correlation is derived from possibly just one quote which pertains to many assets. The method of Buss and Vilkov (2012) is consequently a very different way to proceed which is in practice simpler, because it is easier to obtain vanilla prices for many stocks than to get quote for multivariate derivatives.

Strictly speaking, the definition of the implied volatility requires that the parameter be *implied* strictly, directly and uniquely by the price of the derivative. In the case of the correlation, we believe our method fits this description better.

Chapter 2

Portfolio choice

2.1 Introduction

The topic of portfolio choice and asset allocation has become so vast and the related contributions so numerous that it is impossible to survey exhaustively. Nevertheless, rigourous approaches have all in common a small list components. Namely, they suppose that the agent (or investor) is endowed with the following elements:

1. an **information set**: in a word, **datasets**, which encompass realizations of a data generating process (DGP);
2. **preferences**: often, this is reduced to a utility function, even though this may sometimes be too restrictive, unless it is framed in a very general way (e.g., when investors have style preferences which supersede pure performance maximization);
3. **beliefs**: they define how the agent perceives the world, especially in terms of the data generating process;
4. **constraints**: they are sometimes given by practical contractual obligations and encompass regional and industry weights, factor exposure, leverage and turnover bounds, etc. Asset liquidity is also a frequent concern.

Most design choices in papers on portfolio choice[1] could be integrated into one of these components.

This chapter tackles some of these issues. In Section 2.2, we focus on one particular type of portfolio policy, the **minimum variance** allocation, whereby the agent aims to minimize the risk (proxied by the variance) of a fully invested portfolio. This part covers several facets of this topics: diversification, estimation, and out-of-sample performance. In Section 2.3, we shift the focus to **characteristics-based investing**, whereby agents form their decisions according to assets' (firms', in our setting) characteristics (size, performance, risk, etc.). Lastly, the final part of the chapter (Section 2.4) is dedicated to an unusual challenge: building equity portfolios which behave as much as possible like bond portfolios.

[1]Such as: estimation methods or the choice of a forecasting technique for returns, or the type of model that generates the data.

2.2 Diversified minimum variance

One cornerstone principle of modern portfolio theory is mean-variance allocation, which goes back at least to Markowitz (1952). The formulation is incredibly simple as it seeks to maximize, with respect to a vector of weights $\boldsymbol{w}$ the quadratic utility $\boldsymbol{w}'\boldsymbol{\mu} - \gamma\boldsymbol{w}'\boldsymbol{\Sigma}\,\boldsymbol{w}$, where γ is the risk aversion parameter, $\boldsymbol{\mu}$ is the vectors of average (expected) returns and $\boldsymbol{\Sigma}$ is the covariance matrix of asset returns. Under the budget constraint $\boldsymbol{w}'\mathbf{1} = 1$, the solution is $\boldsymbol{w}^* \propto \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, where the scaling factor in the proportionality relationship is chosen to satisfy the constraint. Equivalently, the optimal portfolio maximizes the expected return for a given level of risk, or minimizes the volatility for a fixed level of performance.

There are two main issues with the above expression for $\boldsymbol{w}^*$. First, when assets are overall positively correlation, hedging relationships (see Stevens (1998) and Goto and Xu (2015)) will imply that many assets are shorted, i.e., that they have negative positions, possibly large in magnitude. This is very easy to check in two dimensions: the inverse covariance matrix is scaled by $(1-\rho^2)^{-1/2}$, where ρ is the correlation between the two assets. As it increases to one, positions become extreme.

For technical reasons like leverage constraints, this may not be authorized for the trader or asset manager. Second, the estimation of the two components $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is very hard. The relative impact of both terms is hard to evaluate (though some have tried, see Chopra and Ziemba (1993) for example). In fact, the true crux of the problem is less estimation (which thereby implies some form of stationarity) than **prediction**. Naturally, the out-of-sample performance is maximized only if the inputs in the optimization are those of *realized* returns, which, by definition, no-one knows in advance.

Heuristically, errors in means are very detrimental because a large expected return will usually imply a large positive position, meaning that an error in the wrong direction will prove financially hurtful. Errors in the covariance structure may seem less impactful but the inversion of the matrix can magnify them significantly. Indeed, when N asset are considered, the matrix requires the estimation of N variances plus $N(N-1)/2$ covariances. Sample sizes are usually much smaller. If the investment universe is the S&P500, this makes 125,000 points to estimate. Working with daily points, this would require 500 year of data to match the dimension. But even if this amount of information was available, returns are notoriously time-varying (Chordia and Shivakumar (2002), Ilmanen (2011)), so that past information, especially if it goes back far in time, may not be suited for **predictive** purposes.

One useful shortcut is to be agnostic with respect to average returns, i.e., to take them to be equal across *all* assets.[2] Interestingly, this corresponds to the case where the investor seeks to minimize the variance of the portfolio ($\boldsymbol{w}'\boldsymbol{\Sigma}\,\boldsymbol{w}$), so that optimal weights are proportional to $\boldsymbol{\Sigma}^{-1}\mathbf{1}$. This formulation now only depends on the covariance structure of assets, but again, when correlations are positive, this will

[2]Empirically, this is often found to be quite effective, given how hard expected returns are to estimate. DeMiguel et al. (2009) show that the naive equally-weighted portfolio is a hard benchmark to beat. Assuming equal expected returns and identity covariance matrix leads to uniform weights.

mechanically lead to negative weights. Short selling can be viewed as a sort of bet: the agent expects some stocks to underperform and sells them to obtain leverage to buy more of the other stocks that are predicted to perform better. This implies a strong form of **under-diversification**. Typically, when leverage is high, the inverse Herfindahl index $1/\boldsymbol{w}'\boldsymbol{w}$ is low, which means that the portfolio is not well balanced and diversified (see Goetzmann and Kumar (2008)).

There are several ways to circumvent the obstacle of concentration in weights. One of the reasons why optimal portfolios are leveraged when they are built on sample estimates is the covariance structure of assets. When assets (e.g., stocks) are strictly positively correlated, the optimization scheme will seek to create arbitrages between stocks, which can also be seen as hedging portfolios (see Stevens (1998), Goto and Xu (2015) and Deguest et al. (2018)). In order to reduce the magnitudes of the hedging relationship, it suffices to attenuate the impact of cross-asset correlations. The easiest way to do so is to strengthen the diagonal of the covariance matrix,[3] so that the relative importance of pure covariances is weakened. In the limit, when the diagonal has infinite value, weights are proportional to the inverse of variances and when the agent is agnostic to them, the equally-weighted portfolio is recovered.

There are several links to various streams of the literature that are associated to this idea. The first is the **shrinkage** approach to covariance estimation. Strengthening the covariance with the identity matrix is what Ledoit and Wolf (2004) recommend. More general techniques are highlighted in Ledoit and Wolf (2012). See also Ledoit and Wolf (2017) for their application in mean-variance portfolio choice and Engle et al. (2019) for dynamic versions thereof. A second related idea is the regression approach of Britten-Jones (1999). It is shown that optimal weights can be obtained by regressing returns against unit vectors. Resorting to shrinkage in this case is equivalent to performing ridge (or Tykhonov) regularization on the coefficients of the regression. Finally, the shrinkage can also be interpreted as a mitigation of noise in the principal components of the portfolios (see Pedersen et al. (2021)).

Note that the practical form $(\boldsymbol{\Sigma}+c\boldsymbol{I})^{-1}\mathbf{1}$ does not solve all issues. Indeed, there remains to settle on the free parameter c. As c increases, the portfolio converges to the uniform allocation. Notably, it is shown in Coqueret (2015a) that at some point (for c large enough), all weights in the portfolio are positive. Thus, c allows to precisely fine-tune the leverage of the portfolio, which is convenient in practice.

Another pending caveat is the dynamic evolution of covariance matrices. Indeed, both variances and covariances evolve through time, thus, estimates based on past data are often outdated when implemented for out-of-sample trading. The importance of the resulting estimation errors is also studied theoretically in Coqueret (2015a).

Empirically, minimum variance portfolios fare quite well. Not only do they deliver robustly lower risks, but they also sometimes also outperform. One reason might

[3] The weights now being proportional to $(\boldsymbol{\Sigma}+c\boldsymbol{I})^{-1}\mathbf{1}$ for some $c>0$. A more direct and radical approach is to simply set bounds on weights in the optimizer. While theoretically sub-optimal, this in facts mitigates estimation risk (Jagannathan and Ma (2003)).

be the low-risk anomaly (Ang et al. (2006), Ang et al. (2009), Baker et al. (2011), Frazzini and Pedersen (2014)). In Coqueret (2015a), I find that all specifications of penalized minimum variance portfolios beat the equally-weighted benchmark, which is a competitive yardstick.

2.3 Characteristics-based investing

In the traditional utility-optimizing literature, portfolios are derived from the statistical properties of their returns. While the first two moments are the most often used, higher orders are sometimes exploited as well (Jondeau and Rockinger (2006), Harvey et al. (2010)).

Nevertheless, with the increased amount of data that is available even through traditional providers (Bloomberg, Thomson-Reuters, Capital IQ), it would seem suboptimal not to use larger information sets in the allocation design. It has become more and more evident that many investors build their portfolio based on other criteria, like valuation ratios (Betermier et al. (2017)), or sustainability (Lagerkvist et al. (2020)) for instance. The cross-section of investor preferences is studied or exploited in contributions such as Cronqvist et al. (2015), Betermier et al. (2017), Koijen et al. (2020) and Balasubramaniam et al. (2021).

One early (yet efficient) attempt to automate the investment process by taking firm characteristics into account is the seminal contribution of Brandt et al. (2009). Their idea is to build a portfolio policy that is defined as adjustments to a given benchmark. The adjustments will be linearly driven by firm attributes. In short, the portfolio weights are $\boldsymbol{w} = \bar{\boldsymbol{w}} + \boldsymbol{X\theta}$, where $\bar{\boldsymbol{w}}$ is the benchmark and $\boldsymbol{X\theta}$ the deviations - decomposed into firm characteristics $\boldsymbol{X}$ and the optimized vector $\boldsymbol{\theta}$.

In theory, the framework can adapt to any simple utility function. Typically, the authors chose to work with the CRRA form $u(x) = x^{1-\gamma}/(1-\gamma)$. However, in practice, this does not work. The reason is that if the estimation sample is such that one characteristic is always associated to a positive return, it should have an infinite value in the $\boldsymbol{\theta}$ vector.[4]

This is why in Ammann et al. (2016) we introduce a regularization to prevent extreme positions. In spirit, this correction, is similar to the L^2 (ridge regularization) discussed in Section 2.2. Simply put, we work with a quadratic utility function, and restrict the portfolio weights so that their L^2 norms $\boldsymbol{w}'\boldsymbol{w}$ is smaller than some value. The combination of quadratic utility and L^2 constraint allows to obtain a quasi closed form for the solution of the program. Numerically, this enables rapid backtests, which is convenient given the number of degrees of freedom in the implementation process.

Notably, one design choice is the set of characteristics to be fed to the optimization. In Ammann et al. (2016), all combinations of two and three characteristics are tested, and reveal the discrepancies in their relevance with respect to their ability

[4]This has been confirmed by at least two independent studies: Gehrig et al. (2018) and one unpublished practitioner paper. Originally, we wanted to include a proof of the error of the original RFS paper in our article, but a reviewer suggested that we remove it.

fo explain and predict the cross-section of returns. In addition, a model in which *all* firm attributes are used is not found to add any value. Thus, picking a small number of appropriate variables is key in the construction of the portfolio weights.

Nevertheless, our results in Ammann et al. (2016) suggest that it is possible to make characteristics-based policies work, if and only if, suitable implementation choices are made. Whether or not these choices can be guessed without forward looking bias is not obvious. The degrees of freedom are the following: the combination of characteristics, the amount of leverage, the risk aversion level and the sample size (shorter samples are more reactive, but carry less information). The best results are obtained for low risk aversion, high leverage and short estimation samples.

To close this section, I wish to briefly mention the topic of factor timing, which is the subject of an unpublished working paper (Amenc et al. (2015)). Because each asset pricing factor (e.g., size, value, momentum) has a strongly time-varying realized return (Ilmanen et al. (2021)), expected return, or risk premium, it is the holy Grail of any investor to be able to determine when it will perform well, or badly. There are many ways to try to predict good and bad time for factors, see, e.g., Bender et al. (2018), Dichtl et al. (2019), and Haddad et al. (2020). In Amenc et al. (2015), we proceeded more in the spirit of the former, i.e., by basing our analysis on exogenous macro-economic variables.

The core of the empirical study is simple: we identify two regimes for each macro variable (high versus low), for instance: high GDP, inflation, credit spread, versus their low counterparts. We then compute the average returns of factors during the low versus high periods for each macro variable.

Our results first show that for some factors (size and momentum), a timing strategy makes sense, notably from an economic point of view, but that for some other factors (value and low risk), the macroeconomic conditioning is not able to discriminate between good and bad periods, at least for the variables we have studied.

The results for size (SMB, Small minus Big) suggest that the factor performs well when the credit spread is small, that is, when smaller (riskier) firms have access to decent, relatively low borrowing rate. This boosts their growth opportunities which may partly explain why the factor behaves well. Another important variable is aggregate economic growth: the size factor performs well in good economic times, however, during crises, small firms are more vulnerable and the profitability of the factor is in jeopardy.

For the momentum factor, our findings corroborate the momentum crash phenomenon (Barroso and Santa-Clara (2015), Daniel and Moskowitz (2016)). The explanation is simple: after an aggregate market crash, the stocks with the most upward potential are those that suffered the most. Thus, shorting these assets is a very bad idea, which leads to momentum poorly performing *after* market downturns. All of these results are of course very valuable for investors who which to build portfolios in which the building blocks are these factors, or proxies thereof. Typically, most investors are long only and can only afford to allocate across thematic ETFs (or, if they are sophisticated enough, build their own factors from individual stocks).

Two factors under scrutiny were found to be very hard to time: value (related to the book-to-market or price-to-book ratio) and volatility. The former has recently been under scrutiny given its very disappointing performance in the second half of the 2010 decade (Shea and Radatz (2020)).[5] Nevertheless, it is possible that other variables may have timing abilities that we have not identified. For example, Hsu and Chen (2017) document that investor attention is useful to improve low risk investing.

Since the publication of Fama and French (1992), characteristics-based investing has boomed in the investment industry, largely thanks to the diffusion of the notion of asset pricing anomalies and to the creation of so-called *Equity Traded Funds* that allow to allocate to particular investment styles. In addition, techniques have evolved from simple linear schemes to much more complex models, and we refer to Section 3.4 and to Coqueret and Guida (2020a) for more on this subject, as well as to sections 3.4, 4.2 and 4.3.

2.4 Equity portfolios that behave like bonds

In the money management, it is often the case that two major asset classes dominate the composition of portfolios: equities and bonds.[6] This simple dichotomy allows to efficiently adjust the risk of the global portfolio by allocating more to equities to increase risk, or more to bonds to decrease it. At a high level, the split between the two classes gives a coarse estimation of the global risk of the allocation.

Nevertheless, for some particular financial institutions, bonds have one very favorable property: they yield regular cash-flows. These flows are associated to low risks which are evaluated by the credit worthiness of the issuer. The US government (for a sovereign bond) is considered as safe bet, while a start up (for a corporate bond) a risky one.

By nature, bonds are designed to generate recurrent payments from the emitter to the buyer. This is very convenient when one faces, in turn, regular outgoing flows, which is typically the case of pension funds. The latter, because they must provide pension benefits to their retirees, need to ensure that they periodically receive inflows of cash. To them, bonds are a blessing.

However, the issue with bonds is that their return profiles are much below those of stocks: as in most themes in Finance, there is naturally a risk/return trade-off. For instance, in the US market, estimates for the average equity return range between 8% and 12%, depending on the period, on calculation method (arithmetic versus geometric) and on dividend inclusion (see, e.g., Ilmanen (2011)). However, average bond returns mostly lie between 3% and 7%, depending on their risks (rating of

[5]Quoting a 2021 Cliff Asness (CEO of AQR) interview: "*Everyone knows the value strategy has been quite bad for the last 10+ years with the 2018-2020 period essentially a crash*".

[6]Other classes, like real estate, credit derivatives, commodities, cryptocurrencies are either smaller or less liquid, and prominent financial institutions cannot trade large amounts without shifting the market. Currencies are less prone to long-term profits because of their mean-reverting property - though some hedge funds can of course use statistical arbitrage on these markets as well.

the issuer and maturity). Though it may seem somewhat unconventional, we may ask the following question: would it be possible to simply combine the idiosyncratic benefits of the two asset classes - all in only one class? That is craft an association (portfolio) that generates recurring cash-flows while at the same time benefits from relatively high average returns.

If we restrict the universe to bonds, increasing returns would require increasing leverage and shorting some bonds in order to buy more of other bonds that are expected to perform well. This is not impossible, though by no means easy to do, especially on a large scale. Thus, the problem is more likely to be solved with stocks.

There are two key properties of bonds which equities typically do not have and which we have discussed above: they have significantly lower risk and they make *nonrandom* (up to credit risk) payments. Again, these are two very desirable properties for pension funds.

In Coqueret et al. (2017) we propose a solution to the aforementioned problem by investigating two key characteristics of firms:

- stocks with low realized volatility;
- companies that have a history of paying large dividends (with respect to the value of their shares).

The rationale for these two indicators is straightforward. The first is the simplest measure of risk, though of course it has many imperfections. The second, proxied by the realized dividend yield, evaluates how much payoff can be expected not from market fluctuations, but from the fundamental activity of the firm, via its benefits and the payout policy of the management.

In addition, these two variables (volatility and dividend yield) are very stable in time and in the cross-section of stocks. Volatility, by construction is slowly-varying and persistent, and low-vol firms, absent idiosyncratic shocks, remain low risk bets (in the cross-section of assets).

Likewise, firms that distribute dividends are not likely to cut them because shrinking dividends sends a very negative signal to the market. Thus, CEOs are inclined to smooth their dividend policies (Leary and Michaely (2011), Chen et al. (2012)). Consequently, dividend yields, like volatilities, are *rather* stable in time and across stocks. This is very important because this reduces the odds of negative surprises.

Given these stylized properties, a natural step in asset pricing is to construct portfolios sorted on these two characteristics. Volatility-based sorts are already widely documented (Baker et al. (2011)), and many different yet similar versions have been implemented: idiosyncratic volatility (Ang et al. (2009)), market beta (Frazzini and Pedersen (2014)) to list but a few. On the other hand, dividends-based investing is an old idea (Brennan (1970), Naranjo et al. (1998)) and makes sense economically.

In Coqueret et al. (2017), we show that selecting stocks that either have low past volatility or high dividend yields not only reduces volatility (compared to a broad equity market index), but improves the realized Sharpe ratio. For both criteria, it improves the average dividend yield of the portfolio. In addition, if we perform a double sort, whereby we select only stocks with low risk and high yield, the results are magnified. Given that these portfolios are less risky, this means that for a given risk budget, the investor can even invest more in equities and further improve its average performance. In the end, the goal is reached: we are able to build equity portfolios with lower risk, higher payout rates (dividend yields that are almost twice those of the market), and higher or comparable returns.[7]

2.5 Heterogeneous agents models

To close this chapter, I would like to spend a few lines on a research subject that is not mentioned in the present document, but which has accounted for a reasonable part of my research between 2015 and 2018: agent-based modelling (ABM).

There is no unambiguous way to define ABMs, other than economic and computation models that simulate the interactions between agents. This is a very broad premise, and it is obviously highly suitable for finance. Notably, early theoretical contributions in finance simplify the modelling part by assuming a representative agent, and heterogeneous agent models (HAMs) are a natural generalization which hopes to describe financial markets more accurately. The link with portfolio choice is clear: agents dynamically form their allocations based on some preferences and/or beliefs and, at each period, the market clearing mechanism updates the price of assets.

In most, if not all early articles (such as the seminal work of Brock and Hommes (1998)), academics focused on a very limited number of agent types, often between two and four. While they already allow to simulate stylized behaviors and preferences (e.g., fundamental versus noise traders), they remain somewhat limited.

In Coqueret (2017b), I propose a simple model that allows for an arbitrarily large number of agents. However, the agents differ across only two dimensions: the intensity with which they are trend-following (or contrarian) and the amount of past history they use to form their beliefs (i.e., how far back in time they go to estimate returns). There are then two questions that I ask: is this model able to replicate stylized properties of asset returns and in which configuration is the fit with empirical data the best?

In turns out that, indeed, the model generates relatively realistic patterns (tails, autocorrelations, volatility clustering). In addition, the best results are obtained when both trend-followers and contrarians have roughly the same ex-ante weight in the market. Furthermore, long term-driven investors are not enough to yield the best fit: agents that react to short term returns are required to match the empirical distribution of returns efficiently.

An interesting topic in ABMs and HAMs is the convergence of deterministic systems to equilibrium points. Sometimes, this may never occur and the models are degenerate. In Coqueret and Tavin (2019), we tackle this challenge with a fairly

[7] In full transparency, for some parametrizations, the average returns were not significantly different from those of the benchmark.

standard evolutionary model. We study it through the lens of the procedurally consistent equilibrium (PCE) introduced in Anufriev and Bottazzi (2010). The requirement of the latter is that, upon equilibrium, investment decisions and agents' wealth remain constant (i.e., reach a fixed point). This stands in contrast with a large portion of the literature in which equilibria impose that prices or returns be static.

In our paper, the key driver of heterogeneity in assets is the dividend yield. Equilibrium returns depend on this yield, as well on the portfolio allocations and on the equilibrium composition of the market. The most interesting facet is this last point: when the system has converged, who are the investors who have survived and how did market selection operate? It turns out that a very specific criterion decides which agents converge to a nonzero market share. It depends on a particular weighted average dividend yield that is derived from agents' portfolios. Simply put, if all agents have the same wealth, the most realistic configuration implies that the surviving agent is the one with the highest average dividend yield. While this statement is simple, the path to its proof is intricate. Nevertheless, it provides a theoretical grounding to strategies that rely on dividend yields.

Chapter 3

Data science and finance

3.1 Introduction

Finance is a discipline that is fundamentally **data-intense**. Of course, a large proportion of the data generated and handled by financial institution remains private (e.g., banking information). Two fields in which data science has thrived is fraud detection and credit scoring. In practice, models and solutions are only deployed in *private* contexts. Nevertheless, the amount of publicly available data is unfathomable. First, there are simple quotes, many of which can for instance be accessed on Yahoo Finance. Quotes are disclosed for all asset classes, even cryptocurrencies, as long as one is able to use a trading platform's API (e.g. Coinbase or Binance, in the crypto trading space).

Beyond prices, other fields are always available that describe assets. For companies, this is obvious, often comes from regulatory financial statements, and forms the bedrock of factor investing. Bonds can be described by the amount issued, the rating of their issuer (credit risk), or its maturity. Cryptocurrencies are characterized by the volume they experience (often, relative to the total value of their network), their hash rate, and their staked amount.

Most of these items can be accessed for free when scrapping websites. However it is much easier, though possibly expensive to download them from a dedicated data provider, such as Bloomberg, Thomson-Reuters, Standard and Poors, or Capital IQ. These services open the doors to even more data. Time-series are readily obtained, which is often not possible from static figures from websites.

Since the middle of the 2010 decade, specialized firms have started gathering, and/or processing more advanced data, which are referred to as *alternative data*. For instance, satellite imagery can be used to nowcast oil exports from Middle-eastern harbours, or to predict local or aggregate retail sales from occupancy rates in parking slots of supermarkets. Likewise, credit card logs are used to forecast aggregate metrics (sales, earnings, or even macro consumption).

This chapter consists of three parts which share little in common, other than they deal with financial data. The first one is about data generation which is increasingly useful in machine learning, or in contexts in which it is desirable to talk about ML (e.g., fintech startups). The second theme is investor sentiment. Notably, if investor sentiment has one obvious value, it would be some predictive power over markets. I show in Section 3.3 that it's not really the case. Finally, the last section of the chapter deals with the idea of boosting factor investing with nonlinear methods stemming from ML.

3.2 Data generation for fintech start-ups

Nowadays, founders of start-ups in the financial sphere feel somewhat obliged (rightfully or not) to include fancy buzzwords in their marketing pitches, such as "*artificial intelligence*" or "*deep learning*". This section revolves around one such example for a robo-advisory platform.

Robo-advisors seek to automate the allocation process for retail clients (notably for retirement savings). Adopters of such solutions are likely to have an edge, compared to money managers that rely on traditional methods (Uhl and Rohner (2018), D'Acunto et al. (2019)).

One crucial step of the customer process is the definition of the level of risk aversion, which will subsequently determine the composition of the portfolio (equity versus bond repartition mainly). Robo-advisory solutions rely on short questionnaires that seek to zero in on the preferences of the client. Such questions assess the wealth and income, as well as the knowledgeability of the customer on financial markets. In addition, a few questions test the propensity to take risky bets.

One of the core component of robo-advising is the translation of clients' responses into a relevant allocation. Originally, when the pool of client is small, this must be done via a rules-based algorithm (that uses heuristic decision branches "*if this then that*"). With a sufficiently large number of clients (and answers), the model that maps responses to allocations can be swapped to a data-based, supervised, method. This is where the marketing pitch of the start-up could mention AI buzzwords (though, in all fairness, simple supervised algorithms are very weak forms of AI).

The main problem for a start-up is that it has few, if any, clients, so that it is impossible to sell a technology that is based on hypothetical data. This is where a small subfield of data science, namely synthetic data generation, can provide valuable assistance. This is the topic of Coqueret (2017a).

One of the challenges that the paper tackles is the complex nature of the data which comprises numerical, ordinal (ordered categorical), and nominal (unordered categorical) variables. It is often feasible to sample purely numerical multivariate distributions, especially if they have a known density and/or belong to familiar parametric families. It is significantly harder to generate samples with several types of variables, each having its own marginal distribution. Relying on the copula approach is feasible, but not with nominal variables. In addition, copulae either not flexible in high dimensions, or hard to tackle (for example, Oh and Patton (2017) rely on simulation-based estimations of their factor-driven copulae).

The first important step in Coqueret (2017a) is to split the variable types. Because of their non-numerical nature, nominal variables are treated separately, and for each of their value, the distribution of all other variables will be generated. Basically, we will sample sub-distributions, conditionally on the realization of nominal variable. Next, we translate ordinal variables into numbers in the simplest fashion possible, that is, in integers. Of course, this could be extended without any loss of generality to other, more complex mappings.

Finally, we are left with mixtures of variables that are close to continuous (e.g., age, wealth), or very discrete (knowledge of financial markets, on a scale from 1 to 5). The sampling of such laws is far from obvious and we resort to a judicious idea called *NORTA*, see Chen (2001). The premise is to

1. generate Gaussian (multi-)variates;
2. turn them into uniform [0,1] variables, via the Gaussian CDF;
3. use the inverse CDF of all marginal distributions to obtain the desired outcome.

Naturally, there are several technical issues, which are mentioned in the paper. The most important one is the dependence structure. General copulae are too complex to handle (and with limited benefits): their estimation in large dimensions is cumbersome. This is why restraining the dependence to the covariance matrix is convenient. The remaining concern is the alteration of the covariance matrix specified in the first point above by the following two steps. Indeed, after the third step, the final covariance matrix is not the one that the user specified originally.

In the paper, we propose an iterative procedure that converges to the sought covariance matrix, as long as the matrix satisfies some mild technical assumption. This can be done because the link between the covariance provided in step one and the one after step three is known (though it is far from obvious).

Extensive simulations confirm our technique which is used to generate arbitrarily large datasets. Unfortunately, as the purpose is to augment a sample, there is no way (to the best of our knowledge) to determine how well this this in fact achieved.

To close this section, I would like to mention that since 2017, a new tool has emerged to sample synthetic data, or augment it: Generative Adversarial Networks (GANs). They are based on two neural networks, one focused on a prediction task, and one aiming to fool the first (in order to improve its performance). Applications of GAN-based data generation were (and still are) mostly concentrated in the field of computer vision (Alaa et al. (2021)), but financial prospects are blossoming (Marti (2020)).

3.3 The relevance of firm-specific sentiment

As recalled in Section 3.1, the field of finance is data intense. One implication is that it has become easier to try to quantify expectations or sentiment towards particular assets, or the aggregate market in general. While these metrics can be mistaken as views of a representative agent (or wisdom of the crowds) because they combine large panels of opinions, they naturally depend on the methodology used to compute them. Thus, a sentiment indicator from Bloomberg will depend on a proprietary set of techniques and databases, which will of course differ from that of a competitor.

The topic of sentiment has thrived in the asset pricing literature since the seminal work of Baker and Wurgler (2006). There is non unique and unambiguous definition for sentiment, and according to Baker and Wurgler (2007), it can be summarized as "*a belief about future cash flows and investment risks that is not justified by the facts at hand*".

The notion of sentiment is appealing in the financial literature. One reason underpinning this attractiveness for researchers relates to the blossoming of behavior-driven models in asset pricing. Prices are dictated by supply and demand and the latter is likely to depend on some form of sentiment. This explains why, in the recent year, sentiment has often been found to be very influential on financial markets, according to academic contributions (see for instance Table 1 in Coqueret (2020c).

Intuitively, sentiment is valuable for an investor if it can be used to derive or increase utility (i.e., returns). One channel to this purpose would be to make expected returns somewhat easier to predict. Several articles have carried out studies on the sentiment's ability to forecast the aggregate market, thus it was more challenging (and valuable) to look at a more granular, level, namely, on a stock-by-stock basis. This is of course once again more data demanding, but all major data providers nowadays have their firm-specific sentiment metrics.

In the case of Bloomberg, an algorithm scans the entirety of the flow of financial news treated by the company and attempts to classify it as good or bad. Then, for a given time interval (one minute, one hour, or one day), the scores are aggregated over all pieces of news. For instance, in a day with a keynote address, Apple will be associated to thousands of news items. Depending of experts' and analysts' expectations and on the announcements during the keynote, the sentiment will be high (usually, positive), or low (usually, negative). Originally, many of them were tags by humans, but the sheer amount of volume imposes automatic treatment, via machine learning and natural language processing (NLP).

The question we ask in Coqueret (2020c) is: can stock-specific sentiment be used to predict returns at short (one day) to medium horizons (one month)? In short, the answer is: no. We run simple predictive regressions for more than 1,000 US stocks (limited by the availability of sentiment scores on sufficiently long time frames) and find that a mere 7% have significant t-statistics (at the 5% level). If we take more conservative confidence thresholds, the number drops sharply. More precisely, for a single asset, the simple model can be summed up as

$$r_{t,t+\Delta} = a + bs_t + e_{t+\Delta},$$

where $r_{t,t+\Delta}$ is a stock's return between t and $t+\Delta$ and s_t is the time-t value of the sentiment for this stock. We run several batches of tests, typically looking at whether closing and opening prices (and the related returns) may influence the results. The time horizon of returns Δ also does not shift the significance of predictability very much.

In addition to simple "*vanilla*" results, the paper discusses the phenomenon of pockets of predictability introduced by Farmer et al. (2021). This interesting idea posits that, like much anything else in finance, predictability is time-varying. During some periods, predictors will do a good job at forecasting one variable, and in some other periods, not so much. In Coqueret (2020c), we also document the existence of micro-level (firm-specific) pockets of predictability, but we also acknowledge that it is a marginal phenomenon.

Finally, for the sake of completeness, we investigate the reverse relationship, whereby returns would influence future sentiment. The most adequate tool for this task, because of its simplicity and its relevance in assessing causation, is the Granger causality test (Granger (1969)). Empirically, we find that returns are 6 times more likely to Granger cause sentiment than the other way around. While the definition of sentiment in financial markets is often vague, our results shed some doubt on the many optimistic positive recent findings in the literature pertaining to sentiment in a broad sense.

3.4 Machine learning for factor investing

3.4.1 Context and principles

The data-rich environment that we have described in this Section is an amazing playing field for supervised learning (SL). Naively, we define the latter as the discipline that searches models that can be broadly summarized as

$$y = f(x) + e, \tag{3.1}$$

where the nature of all elements above can be quite complex. y is most of the time referred to as the *output*, *dependent variable*, *label*, or *target*, while x is often called the *input*, *independent variable*, *feature*, or *predictor*. The term e is simply the error term, or sometimes the *innovation*. For simplicity, f will denote the model, though in other contexts it can be thought of as the data generating engine.

In most financial applications, y is a column vector, while x can be a matrix. The object e has the same dimension as y. Given y and x, the canonical problem in SL is to find one (for simplicity) function f that will minimize a function of the error vector e. The nature of e will depend on that of y and x. If they are real numbers, then finding f is a regression task which is usually solved by minimizing the L^2 norm of e. If they are categories (or classes), then e will take the form of a boolean (true or false) and the classifying algorithm will optimize on more complex functions of errors (e.g., cross-entropy).

As mentioned earlier, typical applications of SL are credit scoring and fraud detection. For credit scoring, y can be either the binary response (grant the loan or not), or the level of the rate, while x embeds information about the borrower (age, gender, income, credit history, etc.). For fraud detection, y is a simple boolean that marks if a transaction was fraudulent or not, and x gathers the information of the transaction (amount, currency, location, type (check, credit card, online), etc.).

The other obvious field of application of SL lies at the intersection of money management and asset pricing. The purpose then becomes to explain, and, if, possible, predict, asset returns. There are of course many ways to proceed, but what has become the conventional approach builds models in a panel fashion. A single, stock-specific, characteristics-based, model would we written:

$$r_{t+1,n} = f_n(x_{t,n}) + e_{t+1,n},$$

and each asset (indexed by n) would have its own model f_n, which depends on its set of characteristics ($x_{t,n}$). Nevertheless, this is reductive because the cross-section of information is very valuable. Thus, the panel-like specification is:

$$r_{t+1,n} = f(x_{t,n}) + e_{t+1,n}, \tag{3.2}$$

where the model f is common to all assets, and all the information available is used to estimate it. Simply put, $f(x_{t,n}) = \mathbb{E}_t[r_{t+1,n}]$ is the time-t conditional expectation of the return. Note that recently, Kelly et al. (2020) go one step further, as they allow for all the dataset to be used to explain or predict the returns of each individual stock.

When the function f in Equation (3.2) is affine (i.e., the model is linear), then, up to some technical specification (e.g., the inclusion of fixed versus random effects), the model is of plain panel type. The whole purpose of supervised learning in general is to consider *nonlinear* forms for f. Decision trees are one of the most widespread **nonlinear** algorithms in SL, and we cover some applications in the next subsection.

3.4.2 Tree methods... and more

Decision trees (DTs) lie at the intersection of clustering and supervised learning. They are both old (they go back at least to Morgan and Sonquist (1963)) and yet powerful, while remaining easy to interpret. DTs seek to split a sample in subsets such that the dependent variable has the highest homogeneity (e.g., the smallest variance) in these subsets. For instance, in regression trees, when the dependent variables is a real number (and not a categorical variable), the objective function is a weighted sum of variances over all subsets. The splits are performed according to the values of the independent variables. We refer to chapter 9 of Hastie et al. (2009) for the mathematical outline of the process.

This tool is very much indicated for factor investing because it automates the portfolio construction process, from a pure data mining point of view. The information pertaining to the firms is exploited in a highly non-linear fashion to build clusters of assets (stocks) with similar performance (the dependent variables is often taken to be a return, possibly relative to some benchmark, but it can be a Sharpe ratio, or an information ratio).

We illustrate this idea in Figure 3.4.1, taken from Coqueret and Guida (2018). The red numbers are the average monthly returns. At the root (the top) of the tree, the whole sample is represented and it has a 0.7% mean return. One important indicator (predictor) in the sample is the Relative Strength Index computed over the past 3 months (RSI 3M), it is a proxy for the past performance of the stock, and is often used in technical analysis. This variable is the first splitting criterion: stocks

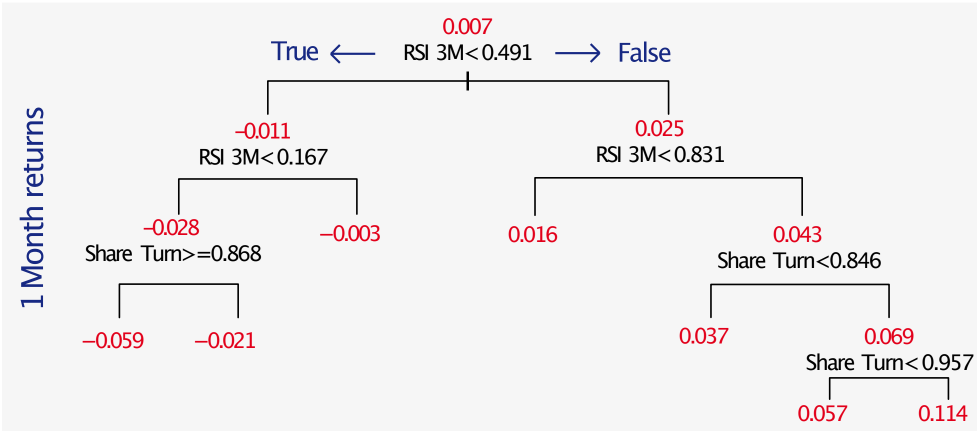

Figure 3.4.1: **A sample decision tree.**

with an RSI below 0.491 make up the first cluster, while the others are gathered in the second cluster.

The process then continues. In the first (left) cluster, the RSI is again the best choice for the split, but the level of the split (0.167) is different. Same thing for the second cluster, for which the splitting level is 0.831. After 2 rounds of clustering, we have obtained 4 groups, and the next splits are based on another predictor, namely the share turnover. The final outcome is a group of 7 clusters, with average monthly returns ranging from -5.9% (left) to 11.4% (right).

Under mild assumptions on the data generating process, it is clear that it should be possible to pursue the splitting iterations until each cluster consists exactly of one observation. This is, so to speak, the curse of DTs: to gain in-sample accuracy, the algorithm must learn precisely and exhaustively what is in the sample. Basically, the algorithm *becomes* the data. This underlines the limits of the generalization ability of DTs, that is, their propensity to generate realistic predictions based on data that they have not been trained on. Indeed, if a very deep DT is presented with an element drawn from a distribution that departs slightly from that of the original data generating process, the resulting error may be substantial.

In order to overcome this issue, statisticians have proposed to rely on ensembles of tree models, whereby several trees should be grown and the final prediction would represent an aggregation of individual predictions. If each tree carries useful information that that is not too correlated with that of the other trees, it is intuitive that a mix of many trees will reduce the magnitude and likelihood of errors.

There are many ways to build ensembles, but 2 main families thereof have emerged:[1]

1. **Random ensembles** are constructed on the principle of bootstrapping. The idea is to grow trees on randomly sampled versions of the data, both in terms

[1]Though popular at the turn of the century, the Adaboost algorithm (Freund and Schapire (1996, 1997)) has mostly been replaced by the two categories we mention below.

of observations (rows) and predictors (columns). The aggregation is then most of the time equally-weighted, meaning that each tree will have the same importance in the overall model. In the case of a tie for a categorical variable, a tie-breaking procedure is defined, e.g., randomly. One central theoretical contribution in this field is Breiman (2001), which shows the virtues of the "*wisdom of the crowds*", especially the tendency of random forests to avoid overfitting, that is, being too close to the sample on which the model is trained, thereby failing to generalize well out-of-sample.

2. **Boosted ensembles** seek to improve the fit of the model whenever a new learner is added to the ensemble. The process is strictly iterative. A first tree is built, and the second one seeks to maximize the fit of the original one by further reducing the in-sample error. However, this is of course the shortest path to only learning the data sample, as a simple tree would do. Therefore, many refinements of boosted trees allows to mitigate the risk of overfitting. They include several penalizations when building the trees that prevent them from sticking too close to the training sample. For instance, a parameter can deter to grow new branches and another one will shrink the values related to terminal leaves which yield the predictions. Also, a learning rate is applied, so that each tree has a marginal importance and leaves room for the next to come.

These two types of tree ensembles have been extremely popular in the recent financial research sphere. We have resorted to them successfully in two practitioner-orientated papers (Guida and Coqueret (2018a,b)). More references, as well as technical details and several examples can be found in the dedicated chapter of Coqueret and Guida (2020a).

Beyond trees, neural networks (NNs) form the other important family of supervised learning tools. In fact, in ML, NNs are now overwhelmingly dominant, if only because the are highly flexible. Different NNs architectures have been shown to crush ML competitions for tasks in computer vision and natural language processing. For problems that rely on tabular data (which is the case in Finance most of the time), NNs do not have a definite edge (Shwartz-Ziv and Armon (2021)). It is notorious that in Kaggle competitions, tree ensembles often fare better, which is why dedicated architectures have been developed (Arık and Pfister (2020), Gorishniy et al. (2021)).

Nevertheless, one drawback of these methods is their lack of transparency. Both neural networks and tree ensembles are complete black boxes because it is impossible to know exactly what the model does, unless one wants to spend hours decomposing its parts (e.g., tree by tree, which would not make much sense). Luckily, in recent years, the literature on ML **interpretability** has blossomed (Molnar (2020), Molnar et al. (2020))

There are many ways to describe interpretation models, but in my opinion, the simplest dichotomy is the one that discriminates between **local** versus **global** models. Global models seek to provide a simple account of the entirety of the model.

One of the simplest way to do this is to determine which variables are the most important and what is their overall impact on the dependent variable. Unsurprisingly, this is exactly what we extract from simple linear models (regressions): the coefficients give the size and direction of the effect, while the t-statistics (or, equivalently, the p-values) give the statistical significance. Researchers are accustomed to these metrics, thus it is intuitive to try to adapt complex techniques to yield similar tools.

Local interpretation models are different in the sense that they do not care about the overall behavior of the model. They are intended to approximate the model in the vicinity of one particular instance. The parallel with the medical field is enlightening. Imagine a doctor has a model that forecasts the likelihood that a patient is subject to an illness. The doctor runs the model and gets a positive outcome. The global patterns of the model does not matter: what matters is why the model predicted the outcome. This comes from the profile of the patient (age, gender, medication, blood pressure, antecedents, etc.) which was given to the algorithm. This is what local models do: they explain, often in linear terms, how the model behaves locally so that the user grasps the predictions for individual observations.

While this is of course very useful in credit scoring, practical applications in the asset management industry are less trivial when the number of assets and predictions is large because it is not feasible (nor necessarily interesting) to analyse a multitude of interpretations pertaining to individual assets. Interpretable ML in asset pricing is mostly a nascent field.

Chapter 4

Present and future work

4.1 Introduction

My recent research encompasses three areas: one is new and the two others are extensions of previously mentioned topics.

The first one is machine learning applied to factor investing. Beyond the simplest applications of supervised learning to stock return prediction, several questions arise which I treat in this chapter:

1. **training**: is it possible to speed up training, and make it possibly more efficient at the same time?
2. **generalization**: it is possible to somehow characterize models that generalize well?
3. **reinforcement learning (RL)**: is RL pertinent for factor investing?

The second (related) domain is asset pricing. Heuristically, if agents form their investment decisions based on firm characteristics, then, mechanically, the latter should drive asset returns. A central question in recent financial economics is: how can we model this dependence?

Finally, the last and new discipline is sustainable investing. Plain honesty requires that I admit that this shift is trend and opportunity driven (I have been asked to work on this topic by a financial institution). Observing the exponential increase in interest and in contribution in this field has boosted my curiosity about it. Moreover, given the long-term consequences for mankind, I felt it was worth exploring in more depth - from an economic and financial point of view at least (the pressure for publication introduces a never-ending bias in my research topics). This has led to a large survey of the literature, an early version of which is Coqueret (2020b), which will become a book by the end of 2021. In addition, I have also started investigating the impact of sustainability criteria in portfolio construction processes. This point is detailed at the end of this section.

4.2 Advanced topics in machine learning & econometrics

4.2.1 Increasing training speed (and possibly model efficiency)

Even though computers have reached nowadays incredible speeds, the ever increasing amounts of data that is available still pushes them to the limit, especially, when it comes to training multiple complex models. Let us take a stylized example, which is representative of the amount of data collected by any reasonable large institutional investor. We consider an investment universe of 3,000 assets (say, stocks, which is already restrictive). For each asset, there are 300 attributes, which range from accounting data, various momentum and risk indicators, to ESG and sentiment scores.[1]. To run a decent backtest, we consider a period of 30 years (360 months). In total, if no points are missing (or if missing data has been imputed), this makes 324 millions points.

Now, in the backtest, models are updated every year, and run on ten years of data at first (possibly with expanding windows subsequently). Ten years of data, means 360,000 rows of data so that a model that does not overfit too much can have up to 100,000 parameters roughly. As a comparison anchor, a simple 3 layer perceptron with 128, 32 and 8 units has roughly 43,000 parameters (assuming 300 predictors and adding biases in every layer). For tree ensembles (random forests or boosted trees), it is also easy to attain these figures by considering large numbers of learners, several thousands for example. Training times for supervised models range from a few minutes to a few days, depending on the sample size, the model complexity, and the computer power. Let us assume an optimistic case of 10 minutes for each annual update, to be carried out on the 20 years beyond the 10 year buffer. This means that it takes roughly three hours to run one single backtest, without tuning the parameters of each model. And this is even without considering exogenous macroeconomic variables that could increase the number of predictors by several orders of magnitude. Moreover, I have not even mentioned how theses signals would be integrated in portfolios, which would take additional coding and running.

Thus, any technique that can help curtailing training times is valuable in ML-based backtests. In full transparency, the idea that lead to the paper Coqueret and Guida (2020b) comes from my co-author, Tony Guida. He tested a simple filtering procedure that removed some observations before sending them to the training phase. Automatically, with smaller datasets, the models take less estimation time, but the gain is not necessarily linear in the sample size.

The filter we apply is the following. Suppose we want to predict (or explain) returns. From an investment standpoint, we are mostly interested by the very high positive returns (which an investor would seek) and the very low negative returns (which an investor would want to avoid). The middle ground, that is, the set of observations with "*average*" returns, is far less appealing and may contain less relevant information. Thus, why not remove it altogether? We propose, at each training stage, to keep only the observations that consist of the top 20% and bottom

[1] Note that at the most granular level, Bloomberg proposes more than 500 indicators for ESG topics *only*!

20% values for the dependent variable.

In the context of regression trees, the paper Coqueret and Guida (2020b) outlines the theoretical consequences of this procedure. In addition, using simulations, it shows that the filtering step is always beneficial[2] to the predictors which have a monotonic link with the dependent variable, which is, in factor investing, a desirable property. Indeed, often, factor portfolios are constructed via simple sorts of high versus low values of the characteristics. Hence, if returns are monotonic in a characteristic, this characteristic may be valuable for the investor.

Our empirical experiment is run with boosted trees. It show that for moderately complex models, shrinking the sample size by a factor two or three is not hurtful. Nevertheless, for shallow models, being able to learn from bigger samples improves the results and our method yields suboptimal results. This makes sense, as a small model trained with limited data seems like a losing combination (though frugal modelling may sometimes be highly efficient). Nevertheless, the core, counterintuitive, message of the paper is that it's possible to slightly improve portfolios by considering smaller training datasets.

In addition to this potential gain in training relevance (the models make more sense, in a way), we document that training times are divided by a factor two to three, which is a modest but tangible gain.

4.2.2 Persistence and accuracy

One important question both in ML and in financial economics is the degree to which what is observed in the past may serve as proxy for what will happen in the future. In ML terms, this is referred to as **generalization**. If a model that is trained on one dataset performs well out-of-sample, i.e., on another (completely different) dataset, it is said to generalize well. This is often relative to some given accepted benchmark (on some pre-specified performance metric).

Just like in statistics, a natural assumption that is often made in the original ML literature is that data and its generation are governed by fixed distributions. It is only under stationary processes that simple theoretical results can be obtained. Heuristically, this makes sense: I can only infer from what I observe. And if distribution shifts occur but I have no information about them, then nothing much can be done.

This is why, in finance, obtaining generalization is very hard, because everything changes all the time: the distribution of returns, the state variables, and the relationships between the two.[3] We refer for instance to Chaieb et al. (2021) and Ilmanen et al. (2021) for a perspective on the time-varying nature of risk premia.

The issue of generalization in non-stationary time-series, depicted in Figure 4.2.1, is therefore challenging. In Coqueret (2020a), I try to find a way to characterize

[2]By beneficial, we mean that, all other things equal, the variable is more likely to be chosen to be the splitting variable at a tree node.

[3]In ML, when the distribution of predictors changes, researchers talk about *covariate shift* (see, e.g., Moreno-Torres et al. (2012)), while when the relationship between predictors and labels change, they refer to *concept drift* (Žliobaitė (2010) and Gama et al. (2014)).

models that perform better and the key notion is **memory**. The rationale is that even if many statistical properties change through time, the changes are not abrupt, but smooth. This smoothness implies some form of latency.

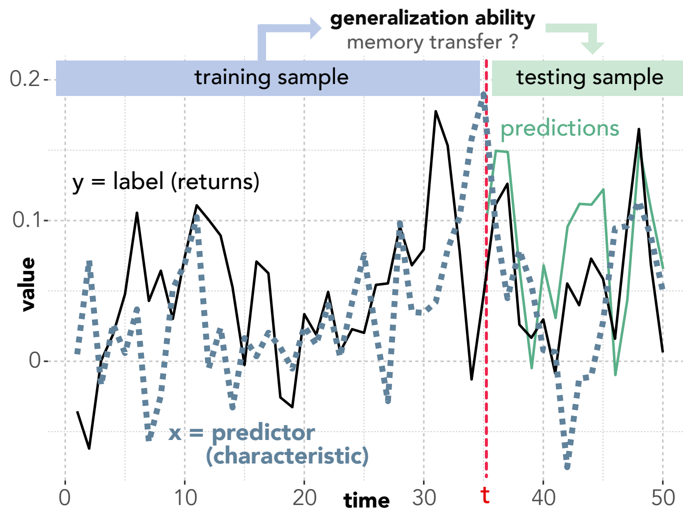

Figure 4.2.1: **Illustration of the concept of chronological generalization for machine learning models**. The model learns from a past sample, but its performance is assessed over future dates. If the accuracy is sustained out-of-sample, the model is said to generalize well.

For the sake of the exposé, let us imagine an agent training a model M_t (at time t). If data samples are deep enough, the model at time $t+1$ will of course share some similarities because there will be a substantial overlap between the two training samples. Loosely speaking, we call *memory* the commonality between M_t and M_{t+1}. Unfortunately, statistical models, in all their generality, are possibly complex tools, so that measuring the distances between one another is not very tractable. For some parametric families (e.g., regressions, neural networks), this could be handled via norms and sums thereof, but for others (tree methods), there is no straightforward way to proceed.

A much more feasible angle is to not consider the models per se, but the data that is fed to them. This is much more easy to handle because, for time-series, we have at our disposal a well-known quantification tool for memory: **autocorrelation**. What is argued in Coqueret (2020a) (and in Coqueret and Deguest (2020)) is that if, on the left-hand side of the equation (3.1) (the label), the time-series are persistent, then the out-of-sample fit are improved, compared to when they are not.

In raw financial terms, this simply means that long term returns are easier to forecast than short term ones. Valkanov (2003) typically writes: "*intuitively, the aggregation of a series into a long-horizon variable is thought to strengthen the signal, while eliminating the noise*". Nevertheless, from a out-of-sample perspective, things are not so simple. Indeed, if I want to build an ML model on annual returns, the values of my predictors must stop one year before present time. This is because the latest (non forward-looking) annual return starts one year from now. This implies that the training data may be somewhat obsolete. There is therefore a **trade-off** between the gain in autocorrelation permitted by long-term returns and the loss of relevance that comes from using possibly out-dated data.

The empirical results in Coqueret (2020a) tend to confirm that there is a net gain overall and that the memory effect is the strongest. More precisely, it is shown that training models on annual returns (rather than on monthly returns[4]) leads to substantially higher returns and Sharpe ratios for portfolios sorted on the related predictions. The effect is more pronounced if the training samples are small but progressively fades as they increase. In addition, we show that the most important predictors are also more persistent than the average input variable. As a confirmation check, we look at *changes* in predictor values (which are memoryless) and find that the resulting performance is, in contrast, very poor.

There remain some important questions. The first one is whether the predictive gain we observe is spurious or not. This question is complex and left for future work, which has started in Coqueret and Deguest (2020). Given the importance of memory, there are promising directions that I envision for factor investing and that have barely been covered by the literature. The first one is the inclusion of past predictor values in the training set. This would substantially increase the amount of independent variables and may boost the risk of overfitting. The second one pertains to recurrent neural networks, which, by construction, are memory-driven. This path has been followed in Cong et al. (2021), but other architectures, such as the one of Dixon (2021) may be more adapted to the task.

Finally, a very important notion that is absent from my work is **causality**. Naturally, it would be a major breakthrough if it were possible to uncover causal relations in asset pricing. My personal feeling is that because of the ever-changing regimes in markets, this is not possible. In Coqueret and Guida (2020a), we estimated a Directed Acyclic Graph (DAG) to try to link firm returns to corporate characteristics. None of the latter entered the graph as predictor (or driver) of the returns. This is a very naive and simplistic example, and of course, the literature in causal inference is abundant, so that this topic should receive more scrutiny. A significant portion relies on the so-called do-calculus of Pearl (2009), which is hard to adapt for financial markets. On a personal level, I am somewhat dubious that this would generate ground-breaking results.

4.2.3 Reinforcement learning and factor investing

Reinforcement learning (RL) is, with supervised and unsupervised learning, the third traditional pillar of machine learning (though, recently, a potential fourth branch, self-supervised learning, has gained traction).

The premise of RL is to learn from **experience** and **rewards**. The agent performs series of actions in an environment. Each action leads to a reward (which can be potentially negative) and to a possible alteration of the environment. Given a series of actions and rewards, the agent progressively learns which actions, in which

[4]Quarterly and semi-annual returns are also tested.

states, yield the best outcomes and thus updates the way he or she acts. If the rules on the global data generating process (environment, rewards, etc.) do not change, then ultimately, if the agent learns well, he or she is able to make enlightened (optimal or close to optimal) decisions.

The domain for which RL has yielded the most impressive results is games: chess, Go and Starcraft. Computers and algorithms are now able to handily beat humans at these games. Indeed, games provide the perfect setting for RL because their rules are relatively fixed (especially for chess and Go), thus, automatic learning is easy, as long as the computer can play enough times (e.g., millions of games).

The transition to finance is not straightforward for at least two reasons. The first reason is the action space. In chess and Go, for any given position (state of the environment), there is a limited amount of moves a player can make. Even if we consider several steps ahead, this number is finite, and in fact, the moves that make sense (i.e., that are not completely stupid) are not that numerous. In a financial environment, the number of actions is infinite.

This is a serious drawback, because most traditional RL algorithms rely on Markov Decision Processes (MDPs) in finite spaces (see, e.g., Sutton and Barto (2018)). Thus, in order to benefit from the RL artillery (Q-Learning and most tabular methods), one must first discretize the action space.

There are several ways to do so in finance, in the framework of portfolio choice. First, one could shrink the investment set to a handful of assets, say two to five for instance. This can make sense for asset allocation, where each block would pertain to a particular asset class. However, in order to obtain discrete decisions, the portfolio weights must belong to a discretized simplex. If we allow for n possible values for k stocks, the number of actions is $(n+k-1)!/(n!(k-1)!)$. For five assets and eleven possible values (from 0% to 100% by tranches of 10%), this makes 1,365 possible actions.

This is, in fact, a pretty big number. If we consider Q-learning, the agent would have to evaluate the average reward for all of these actions, across all possible states of the environment. Naturally, this is a complex task, because it requires to evaluated the performance of all possible portfolio compositions in all economic situations. The number N of cases to consider is therefore the product of the number of these situations times the number of possible actions. To learn, the agent must see (experience) a number of observations (actions and rewards) that is a large multiple of N: this is hardly realistic, at least in large dimensions. In low dimensions, the stratified approach of Tuck et al. (2021) could be translated to RL.

Which leads us to the second major issue: the environment. Again, in a simple game (e.g., tic-tac-toe), it is fairly easy to describe and characterize the environment. This is practically impossible in finance, unless one is willing to make overly simplifying assumptions. Another key assumption in RL that is mostly not verified in finance is that each action of the agent modifies the environment. While this may be true in financial markets, the impact of one agent (unless it is BlackRock) is not likely to significantly move prices. In fact, modern traders try to avoid to impact markets by optimizing their orders.

The issue of finite environments is of course a major barrier for an agent seeking to exploit factor investing. This is because each firm is associated to a possibly large cross-section of indicators which cannot be realistically discretized. Thus, another approach is required if we want to combine RL to asset pricing factors.

In André and Coqueret (2020), we try to circumvent these technical hurdles by resorting to so-called policy gradients. The foundational principle comes from random policies in RL, usually denoted with π. The idea is that, just like for mixed strategies in game theory, the agent will choose an action probabilistically: π defines the odds that he or she should perform this or that action. This is convenient, because the field of distribution is sufficiently large to provide a large palette of choices. The key ingredient in André and Coqueret (2020) is the Dirichlet distribution, which is defined on a simplex, and is thus well suited for portfolio purposes: each value will give the weight of the asset in the portfolio.

The policy gradient method is meant to improve the policy π so that it ends up maximizing an expected reward (e.g., an average return, or a Sharpe ratio). To do so, we are interested in the sensitivity of the reward in changes of the policy, hence the term "*gradient*". This gradient is computed with respect to the parameters of the policy - this imposes to chose a parametric distribution, which is not a major hurdle. The learning process then consists in updating the parameters of the policy via gradient ascent.

The missing piece of the puzzle is the factor part. A Dirichlet distribution of order N has N parameters a_n which drive the average of the marginals (i.e., the average allocation across assets). The idea in André and Coqueret (2020) is to link these parameters to the characteristics of firms. We propose two forms: $a_n = \sum_k \theta^k x_n^k$ and $a_n = \exp\left(\sum_k \theta^k x_n^k\right)$, where the θ^k drive the exposure of the policy to the characteristics x_n^k. The second one is preferable because the parameters of the distribution must be positive, but the first one can be twisted by projecting an unfeasible vector onto the corresponding simplex. In the end, the aim is to compute the gradient with respect to the vector $\boldsymbol{\theta}$.

In our empirical tests on US equities, we included one constant plus 12 firms-specific characteristics stemming from the mainstream asset pricing literature (capitalization, valuation ratio, momentum, etc.). The role of the constant is crucial, as it serves to link the performance to a factor that is common across all stocks.

Our results are both disappointing and quite revealing. They show that the constant strongly dominates all other characteristics, i.e., that the θ^0 value assigned to it dwarfs the others in magnitude. One direct implication of this is that the reinforced portfolios are very close to the equally-weighted ($1/N$) portfolio. This is underwhelming because the technical apparatus that is deployed is substantial, while the outcome is simplistic.

Nevertheless, this also tells a story about the efficiency of financial markets (at low frequencies) and about the pricing ability of characteristics. Basically, since the advent of data providers, equity-traded funds (ETFs), and machine learning, most large money managers take the most common characteristics into account. Thus, whenever an anomaly is detected, it is often rapidly arbitraged away by the markets (see, e.g., Schwert (2003), McLean and Pontiff (2016), Jones and Pomorski (2017), Penasse (2020)).[5]

Overall, this makes it hard to traditional asset pricing variables to keep a predictive ability over returns. This is what is learned by our model. We resort to subsampling (akin to bootstrapping) to continuously update the θ^k which try to capture the link between characteristics and performance (both returns and Sharpe ratios). But, in the end, the ever changing link between them (in time, and in the cross-section of stocks) benefits to the constant only and the $1/N$ portfolio emerges as the optimal choice. This conclusion is not surprising either. The equally-weighted portfolio has been shown to perform well empirically (DeMiguel et al. (2009)), and is theoretically optimal when there is a large amount of ambiguity (Maillet et al. (2015)).

4.3 Characteristics-based asset pricing

Investors, when crafting their decisions, rely on some informational sets. In the equity investment space, as discussed above, a lot of information can be gathered at the firm level (e.g., from accounting data, price data, risk measures, sentiment, etc.). This level of granularity improves decision making process, compared to resorting to macro-economic indicators only.

The idea of characteristics-based portfolios is as old as value investing (Graham et al. (1934)), but has experienced a golden age since the seminal contribution of Brandt et al. (2009) for its systematic implementation (see Section 2.3) and in parallel to the rise of factor investing (see Barberis and Shleifer (2003) for an early theoretical contribution).

Nowadays, with the amount of data that is available, as well as with the computational power of GPUs, TPUs and parallel cores, systematic methods have emerged that seek to mine large datasets so as to extract patterns that link firm attributes to future returns (see, e.g., Suh (2019), Freyberger et al. (2020), Gu et al. (2020), Kim et al. (2021a,b), Windmüller (2021)). This is of course incredibly valuable from a portfolio perspective, if and only if, the links between the characteristics and the cross-section of returns remains stable in time - which is in fact hard to obtain.

More generally, characteristics can be and are used in asset pricing. Heuristically, if agents use them for allocation purposes, then it is only logical that returns should also depend on these characteristics. Naturally, there are many ways to incorporate firm attributes into asset pricing models. A recent trend is to stick to the traditional factor specification, whereby returns are explained by their exposure to a finite set of K factors: $r_{t+1,n} = a_n + \sum_{k=1}^{K} b_{t,n} f_{t+1}^k + e_{t+1,n}$. Note that depending on the task

[5] There is in fact a heated debate on whether or not this may be due to p-hacking and data mining in academic publications in the first place. Journals' publication bias towards positive results is a strong incentive for researchers to reveal only the most promising of their empirical findings. Some defend this position (Harvey (2017), Chen and Velikov (2020), Falck et al. (2021), Harvey and Liu (2021)), while some contest it (Müller and Schmickler (2020), Jensen et al. (2021)). Other argue that it's true, but mostly marginal (Chen and Zimmermann (2020), Chen (2021), Chen and Zimmermann (2021)).

(explanation versus prediction), the time indices may be shifted. The trick is then to assume that the stock-specific factor loadings depend on firm attributes, as in the linear shape $b_{t,n} = a_n + \sum_{m=1}^{M} \beta_n^m x_{t,n}^m + \varepsilon_{t,n}$[6] of Kelly et al. (2019) (see also Cosemans et al. (2016), Dittmar and Lundblad (2017), Connor et al. (2021), and Ge et al. (2021)).[7] Nevertheless, this approach still leaves one important degree of freedom: the choice of the underlying factors f_{t+1}^k. In the asset pricing literature, they can be latent, i.e., defined as an optimal combination of existing returns (see, e.g., Chamberlain (1983), Connor and Korajczyk (1993), Bai and Ng (2002), Kelly et al. (2019), and Lettau and Pelger (2020a,b)), or explicit (either via traditional macro-economic variables, or via deterministic construction (e.g., SMB or HML in Fama and French (1993) or WML from Jegadeesh and Titman (1993))).

An old and simple idea, originating from Daniel and Titman (1997), is that the performance of firms may simply be driven by their characteristics more than by their exposures to some asset pricing factors. In fact, recently, a few studies have shown that regressing returns on characteristics is more efficient (in terms of out-of-sample explanatory power) than regressing them on traditional factors (we refer for instance to Hou et al. (2011), Goyal and Jegadeesh (2018), Chordia et al. (2019), Fama and French (2020) and Raponi et al. (2020)).

In an ongoing research project (much inspired by the incredible paper by Koijen and Yogo (2019)), I show that it is possible to obtain characteristics-driven returns from a partial equilibrium. The only requirement is that agents in the model have a demand d for assets that have two components: one that is a function of the characteristics, say, c, and one that is a multiple of the log-price of the asset (with a negative slope). Indeed, when equated to some exogenous supply s, this gives $d = c - b\log(p) = s$, so that the price is simply $p = \exp((c-s)/b)$. If quantities are indexed by time, taking the log-return yields the sought form.

Additionally, we study the conditions under which asset pricing anomalies (based on portfolio sorts) can occur. Our result reveal that some interactions between characteristics may matter theoretically,[8] but not very much empirically. Moreover, panel regressions reveal that a small set of traditional characteristics can only explain a small portion of anomalies.

4.4 Sustainable equity investing

Socially responsible investing (SRI) has likely been the number one topic in the investment community in 2020-2021. There are at least two possible direct reasons are that. First, green funds and assets have performed very well lately, especially in the aftermath of (the beginning of) the Covid-19 pandemic, i.e., subsequently to the February-March 2020 crash. This makes them naturally very attractive for investors. Relatedly and secondly, flows have poured in sustainable funds, either because of

[6] Here, $x_{t,n}^m$ denotes the time-t m^{th} characteristic of stock n.

[7] Giglio et al. (2021) also argue that stocks differ in their exposure to asset pricing factors.

[8] Interactions betweeh characteristics are investigated in different ways in Müller and Schmickler (2020) and Ross (2020) for instance.

good performance (cf supra), or because the broader public, acknowledging the urgence of climate change, has progressively shifted its savings towards investment that are seemingly less hurtful for the long term perspectives of the planet.

Coming with a strong portfolio choice background, my first inclination was to investigate the link with performance. The natural question should be: how hurtful can SRI be? This is simply because there should be a cost attached to imposing constraints on what would otherwise be optimal portfolios. However, after reading a few dozen articles on the topic, it appeared as if in fact, SRI would be an opportunity. In hindsight, I am afraid that many researchers, willing to sell a nice story about the so-called "*doing well by doing good*" motto, reported overly optimistic results.

Indeed, after having read hundreds of papers on this topic, my meta-view (see Coqueret (2020b)) is that sustainability is in fact not priced very much. This is the middle ground between the two extreme positions, but it is in fact very good news because it means that being green is not costly, or at least too costly.

One important topic in this field is the shifting of investor preferences. Several recent natural experiments (Hartzmark and Sussman (2019), Lagerkvist et al. (2020), Chew and Li (2021)) document the prevalence of sustainability criteria in agents' choices. Unfortunately, the survey of Stroebel and Wurgler (2021) show that experts believe that markets underestimate risks related to climates change.

The recent push towards ESG (Environmental, Social and Governance) disclosure has provided investors with a large palette of new metrics. The good news is that this provides new characteristics, i.e., new predictors for researchers (see section 4.3 above).

The bad news is that there are many inconsistencies across raters and data providers (Berg et al. (2020), Dimson et al. (2020)), which hinders robustness of conclusions. On top of that, the complexity of the data field complicates the understanding of what investors genuinely do. For instance, it is easy to obtain firms' scope 1 and 2 emissions, but it is notorious that they account for often less than half of the emissions of the business line.[9] Given that there is no regulatory requirement on the disclosure of scope 3 emissions, their evaluation is very blurry. Consequently, investors who truly seek to have low carbon portfolios to minimize the impact of their holdings remain somewhat in the dark. This is true well beyond emissions on all three pillars of the ESG trichotomy. The impact of rating disagreement on asset prices has recently been studied in Avramov et al. (2021) and Christensen et al. (2021).

More generally, there is a lot of contradicting evidence on whether or not sustainability is priced, that is, if there is such thing as an ESG factor. Some contributions argue that ESG is positively priced, others that it is negatively priced. Some authors find that it is not priced at all and some show that *it depends* (on, the pillar, on time, on sector, on country, etc.). Quite conveniently, this question relates deeply to Sections 3.4 and 4.4 which cover characteristics-based investing. Indeed, ESG now provides the asset pricing literature with new families of characteristics. My

[9] There are of course sectorial discrepancies but this phenomenon is acute in the industry, as well as in the energy and banking sectors.

own opinion is that ESG is not priced, which can be interpreted in two ways. The pessimistic take is that investors are not incentivized to favor sustainability. The optimistic view considers that, since ESG is at the same time not costly, there is no reason why asset managers and owners should not indulge in green investments.

Finally, the last and most important topic is civilizational: what is the role of financial markets in humanity's handling of global warming? This is a very difficult questions for at least two reasons. The first one is technical: since the DICE model of Nordhaus and Boyer (2000), climate-based economic models have become increasingly complex and harder to estimate. Some economists are even quite critical about these models and their empirical results (Keen (2020)). A second obstacle is that long term effects depend on policy adjustments that are highly unpredictable. Carbon taxes around the world remain small scale initiatives, but, if temperatures continue to rise, they are bound to become more and more binding. Likewise, mandatory disclosure of carbon emissions is likely to be useful in the long run, but the timing of such requirements remains uncertain.

Conclusion

My post-doctoral research has blossomed in several directions. The dawn of my professional career has led me to abandon the original field that I was working on, which was the pricing of complex derivative products by means of sophisticated tools in stochastic calculus. My last contributions on the topic revolve around model risk and the definition of implied volatility.

I have very much enjoyed switching to the more optimization-based facet of finance that is portfolio choice. I feel it is somewhat easier to apply simple theories on market data because the latter is effortlessly downloadable from public sources. In contrast, interesting datasets of derivative quotes are relatively expensive, and most of the numerical pricing that I have done was checked with simulations and not in any circumstance derived from actual market quotes. However, after a few years on this topic, I felt that portfolio optimization and the related challenges of estimation were no longer a field were I could incrementally learn a lot.

Consequently, in the past six years, my focus has mainly been on the interactions between data science and market finance. Given the amount of data that any data vendor can provide, this is a fertile ground for empirical work. One intriguing field is the overlap between machine learning and asset pricing, whereby researchers resort to supervised learning to automate the detection of patterns between returns and large sets of predictors.

While intellectual gratifying at first, this task progressively becomes frustrating because, in spite of all the positive results published in journals (and the marketing deployed by money managing companies), ML falls short of its early promises.[10] My conjecture to why that may be the case is that, nowadays, the numbers of participants on the market who trade based on information flows is such that any inefficiency is arbitraged away very rapidly. ML models have been developed for high-frequency purposes, but as I have not studied them, I do not have a solid opinion on this domain.

The prediction of returns, can be tackled via at least two complementary disciplines. The first route is the one I originally took and pertains to statistics and econometrics. The second route is one that I work on more intensely: asset pricing. I enjoy that most theoretical models have to be grounded on economic reasonings.

[10] I would venture to say that it is true for almost all fields in which ML/AI has been predicted to disrupt an industry. One typical example is autonomous vehicles. I remember telling my students in 2017 that driverless cars would undoubtedly hits the streets in 2020. Needless to say, that was a poor forecast. Progress has been made for sure, but true revolutions have yet to occur.

Equilibrium models which have closed forms are often quite elegant, which is why they are at the core of one of my current papers.

I am nevertheless increasingly preoccupied by the risk of p-hacking in financial economics. The presentation (and article) Harvey (2017) was eye-opening, but I am afraid it only scratches the surface of the problem. The more I read papers in top journals and the more I wonder why authors complicate their empirical protocols to the point where it takes hours to figure them out. In fact, whereas articles in top journals could be very short in the middle of the XXth century, nowadays papers occasionally exceed 80-100 pages (see Card and DellaVigna (2014) for more details, as well as the statement in the following footnote[11], which I would like to keep archived here. It seems that given the demand to publish in top journals (in part because of the consequences on tenure), the barrier to entry has been placed on paper complexity, while, in my opinion, it should be on paper originality, ambition and reproducibility (a topic on which a lot of progress has been made). This is a subject that I am currently reflecting on, even though it is a tough one to tackle.

Lastly, for many reasons, I have gained interest in sustainable investing. After one year surveying this incredibly large domain, I feel that I know more, but that, at the same time, I will never be knowledgeable enough. I plan to complete the 3-4 projects that I have currently (including one book chapter) and see if I come up with new ideas or opportunities. To be honest, like most academics, research ideas is not what I lack the most.

[11]**Statement from the editors of Econometrica, Quantitative Economics, and Theoretical Economics**: Published papers in many of the economics journals have gotten much longer over the years. The Econometric Society journals Econometrica, Theoretical Economics, and Quantitative Economics, are no exception. The average length of a paper in Econometrica was about 12 pages in the 1970s, and is now about 36 pages, not including online appendices, with many papers over 50 pages. Some of this may be due to changes in technology, such as the increased ability to estimate complex models using large data sets, or the greater technical difficulty in some areas in theory or econometrics, but probably not all of it.

These trends are a concern to us. Although Econometrica now publishes considerably more pages than in the 1970s, it publishes fewer papers, while during the same period the number of submissions has gone up substantially. Longer papers are hard on the readers and the reviewers. They take longer to review, and often require more rounds of revisions. Also, in our view it is more challenging to maintain a high quality of writing (of both text and mathematical proofs) when the paper is very long.

We therefore will try to stop, and in fact try to reverse the trend and bring down the paper length (including online appendices). This will take effort on our part, but also requires a change in attitude from authors and reviewers. We encourage authors not to think of a typical Econometrica/QE/TE paper as a 30-40 page paper, but rather aim for a 20-30 page paper We will encourage reviewers to be restrained in their requests for additional analyses and results that will increase the paper length. On our part we will hold a harder line on paper length, including sending back submissions before having them reviewed if they are excessively long, and in revisions set stricter page limits. We are not imposing an overall page limit: the appropriate length may differ substantially between theory, econometrics, and empirical papers and also varies by content. We intend to do more than simply move material from the paper to the online appendices. We aim to reduce the overall length of the paper, including those appendices.

Guido Imbens (Editor Econometrica)

Ran Spiegler (Editor Theoretical Economics)

Chris Taber (Editor Quantitative Economics)

Bibliography

Achdou, Y. and Pironneau, O. (2005). *Computational methods for option pricing*. SIAM.

Ackerer, D. and Filipović, D. (2020). Option pricing with orthogonal polynomial expansions. *Mathematical Finance*, 30(1):47–84.

Aguilar, J.-P. and Kirkby, J. L. (2021). Closed-form option pricing for exponential Lévy models: a residue approach. *SSRN Working Paper*, 3823337.

Alaa, A. M., van Breugel, B., Saveliev, E., and van der Schaar, M. (2021). How faithful is your synthetic data? Sample-level metrics for evaluating and auditing generative models. *arXiv Preprint*, (2102.08921).

Amenc, N., Coqueret, G., and Martellini, L. (2015). Active allocation to smart factor indices. *EDHEC-Risk Institute Working Paper*.

Ammann, M., Coqueret, G., and Schade, J.-P. (2016). Characteristics-based portfolio choice with leverage constraints. *Journal of Banking & Finance*, 70:23–37.

André, E. and Coqueret, G. (2020). Dirichlet policies for reinforced factor portfolios. *SSRN Working Paper*, 3726714.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1):259–299.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2009). High idiosyncratic volatility and low returns: International and further US evidence. *Journal of Financial Economics*, 91(1):1–23.

Anufriev, M. and Bottazzi, G. (2010). Market equilibria under procedural rationality. *Journal of Mathematical Economics*, 46(6):1140–1172.

Arık, S. O. and Pfister, T. (2020). Tabnet: Attentive interpretable tabular learning. *arXiv Preprint*, (1908.07442).

Asmussen, S. and Rosiński, J. (2001). Approximations of small jumps of lévy processes with a view towards simulation. *Journal of Applied Probability*, pages 482–493.

Avramov, D., Cheng, S., Lioui, A., and Tarelli, A. (2021). Sustainable investing with ESG rating uncertainty. *Journal of Financial Economics*, Forthcoming.

Bachelier, L. (1900). Théorie de la spéculation. In *Annales scientifiques de l'École normale supérieure*, volume 17, pages 21–86.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Baker, M., Bradley, B., and Wurgler, J. (2011). Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly. *Financial Analysts Journal*, 67(1):40–54.

Baker, M. and Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance*, 61(4):1645–1680.

Baker, M. and Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, 21(2):129–152.

Balasubramaniam, V., Campbell, J. Y., Ramadorai, T., and Ranish, B. (2021). Who owns what? A factor model for direct stockholding. *SSRN Working Paper*, 3795521.

Barberis, N. and Shleifer, A. (2003). Style investing. *Journal of Financial Economics*, 68(2):161–199.

Barroso, P. and Santa-Clara, P. (2015). Momentum has its moments. *Journal of Financial Economics*, 116(1):111–120.

Bateman, H. (1953). *Higher transcendental functions [volumes i-iii]*, volume 1. McGraw-Hill Book Company.

Bender, J., Sun, X., Thomas, R., and Zdorovtsov, V. (2018). The promises and pitfalls of factor timing. *Journal of Portfolio Management*, 44(4):79–92.

Berg, F., Kölbel, J. F., and Rigobon, R. (2020). Aggregate confusion: The divergence of ESG ratings. *SSRN Working Paper*, 3438533.

Betermier, S., Calvet, L. E., and Sodini, P. (2017). Who are the value and growth investors? *Journal of Finance*, 72(1):5–46.

Bianchi, M. L., Rachev, S. T., Kim, Y. S., and Fabozzi, F. J. (2010). Tempered stable distributions and processes in finance: numerical analysis. In *Mathematical and statistical methods for actuarial sciences and finance*, pages 33–42. Springer.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654.

Brandt, M. W., Santa-Clara, P., and Valkanov, R. (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies*, 22(9):3411–3447.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Brennan, M. J. (1970). Taxes, market valuation and corporate financial policy. *National Tax Journal*, 23(4):417–427.

Britten-Jones, M. (1999). The sampling error in estimates of mean-variance efficient portfolio weights. *Journal of Finance*, 54(2):655–671.

Brock, W. A. and Hommes, C. H. (1998). Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic dynamics and Control*, 22(8-9):1235–1274.

Buss, A. and Vilkov, G. (2012). Measuring equity risk with option-implied correlations. *Review of Financial Studies*, 25(10):3113–3140.

Card, D. and DellaVigna, S. (2014). Page limits on economics articles: Evidence from two journals. *Journal of Economic Perspectives*, 28(3):149–68.

Carr, P. and Bowie, J. (1994). Static simplicity. *Risk*, 7(8):45–50.

Carr, P., Geman, H., Madan, D. B., and Yor, M. (2002). The fine structure of asset returns: An empirical investigation. *Journal of Business*, 75(2):305–332.

Carr, P. and Madan, D. (1999). Option valuation using the fast fourier transform. *Journal of Computational Finance*, 2(4):61–73.

Carr, P. and Wu, L. (2003). The finite moment log stable process and option pricing. *Journal of Finance*, 58(2):753–777.

Chaieb, I., Langlois, H., and Scaillet, O. (2021). Factors and risk premia in individual international stock returns. *Journal of Financial Economics*, 141(2):669–692.

Chamberlain, G. (1983). Funds, factors, and diversification in arbitrage pricing models. *Econometrica*, 51(5):1305–1323.

Chaumont, L. et al. (2013). On the law of the supremum of lévy processes. *Annals of Probability*, 41(3A):1191–1217.

Chen, A. Y. (2021). The limits of p-hacking: Some thought experiments. *Journal of Finance*, Forthcoming.

Chen, A. Y. and Velikov, M. (2020). Zeroing in on the expected returns of anomalies. *SSRN Working Paper*, 3073681.

Chen, A. Y. and Zimmermann, T. (2020). Publication bias and the cross-section of stock returns. *Review of Asset Pricing Studies*, 10(2):249–289.

Chen, A. Y. and Zimmermann, T. (2021). Open source cross-sectional asset pricing. *Critical Finance Review*, Forthcoming.

Chen, H. (2001). Initialization for NORTA: Generation of random vectors with specified marginals and correlations. *INFORMS Journal on Computing*, 13(4):312–331.

Chen, L., Da, Z., and Priestley, R. (2012). Dividend smoothing and predictability. *Management Science*, 58(10):1834–1853.

Chew, S. H. and Li, K. K. (2021). The moral investor: Sin stock aversion and virtue stock affinity. *SSRN Working Paper*, 3773971.

Chopra, V. K. and Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19(2):6.

Chordia, T., Goyal, A., and Shanken, J. A. (2019). Cross-sectional asset pricing with individual stocks: betas versus characteristics. *SSRN Working Paper*, 2549578.

Chordia, T. and Shivakumar, L. (2002). Momentum, business cycle, and time-varying expected returns. *Journal of Finance*, 57(2):985–1019.

Christensen, D. M., Serafeim, G., and Sikochi, S. (2021). Why is corporate virtue in the eye of the beholder? The case of ESG ratings. *Accounting Review*, Forthcoming.

Cong, L. W., Tang, K., Wang, J., and Zhang, Y. (2021). Deep sequence modeling: Development and applications in asset pricing. *Journal of Financial Data Science*, 3(1):28–42.

Connor, G. and Korajczyk, R. A. (1993). A test for the number of factors in an approximate factor model. *Journal of Finance*, 48(4):1263–1291.

Connor, G., Li, S., and Linton, O. B. (2021). A dynamic semiparametric characteristics-based model for optimal portfolio selection. *SSRN Working Paper*, 3803193.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236.

Cont, R. and Tankov, P. (2003). *Financial modelling with jump processes*. CRC Press.

Cont, R. and Voltchkova, E. (2005). A finite difference scheme for option pricing in jump diffusion and exponential Lévy models. *SIAM Journal on Numerical Analysis*, 43(4):1596–1626.

Coqueret, G. (2015a). Diversified minimum-variance portfolios. *Annals of Finance*, 11(2):221–241.

Coqueret, G. (2015b). On the supremum of the spectrally negative stable process with drift. *Statistics & Probability Letters*, 107:333–340.

Coqueret, G. (2017a). Approximate norta simulations for virtual sample generation. *Expert Systems with Applications*, 73:69–81.

Coqueret, G. (2017b). Empirical properties of a heterogeneous agent model in large dimensions. *Journal of Economic Dynamics and Control*, 77:180–201.

Coqueret, G. (2020a). Persistence in factor-based supervised learning models. *SSRN Working Paper*, 3588261.

Coqueret, G. (2020b). Perspectives in ESG equity investing. *SSRN Working Paper*, 3715753.

Coqueret, G. (2020c). Stock-specific sentiment and return predictability. *Quantitative Finance*, 20(9):1531–1551.

Coqueret, G. and Deguest, R. (2020). Predictive regressions: A machine learning perspective. *SSRN Working Paper*, 3709412.

Coqueret, G. and Guida, T. (2018). Stock returns and the cross-section of characteristics: A tree-based approach. *SSRN Working Paper*, 3169773.

Coqueret, G. and Guida, T. (2020a). *Machine Learning for Factor Investing: R Version*. CRC Press.

Coqueret, G. and Guida, T. (2020b). Training trees on tails with applications to portfolio choice. *Annals of Operations Research*, 288:181–221.

Coqueret, G., Martellini, L., and Milhau, V. (2017). Equity portfolios with improved liability-hedging benefits. *Journal of Portfolio Management*, 43(2):37–49.

Coqueret, G. and Tavin, B. (2016). An investigation of model risk in a market with jumps and stochastic volatility. *European Journal of Operational Research*, 253(3):648–658.

Coqueret, G. and Tavin, B. (2019). Procedural rationality, asset heterogeneity and market selection. *Journal of Mathematical Economics*, 82:125–149.

Coqueret, G. and Tavin, B. (2020). A note on implied correlation for bivariate contracts. *Economics Bulletin*, 40(2):1388–1396.

Cosemans, M., Frehen, R., Schotman, P. C., and Bauer, R. (2016). Estimating security betas using prior information based on firm fundamentals. *Review of Financial Studies*, 29(4):1072–1112.

Cronqvist, H., Siegel, S., and Yu, F. (2015). Value versus growth investing: Why do different investors have different styles? *Journal of Financial Economics*, 117(2):333–349.

D'Acunto, F., Prabhala, N., and Rossi, A. G. (2019). The promises and pitfalls of robo-advising. *Review of Financial Studies*, 32(5):1983–2020.

Daniel, K. and Moskowitz, T. J. (2016). Momentum crashes. *Journal of Financial economics*, 122(2):221–247.

Daniel, K. and Titman, S. (1997). Evidence on the characteristics of cross sectional variation in stock returns. *Journal of Finance*, 52(1):1–33.

Deguest, R., Martellini, L., and Milhau, V. (2018). A reinterpretation of the optimal demand for risky assets in fund separation theorems. *Management Science*, 64(9):4333–4347.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, 22(5):1915–1953.

Di Nunno, G., Øksendal, B. K., and Proske, F. (2009). *Malliavin calculus for Lévy processes with applications to finance*, volume 2. Springer.

Dichtl, H., Drobetz, W., Lohre, H., Rother, C., and Vosskamp, P. (2019). Optimal timing and tilting of equity factors. *Financial Analysts Journal*, 75(4):84–102.

Dimson, E., Marsh, P., and Staunton, M. (2020). Divergent ESG ratings. *Journal of Portfolio Management*, 47(1):75–87.

Dittmar, R. F. and Lundblad, C. T. (2017). Firm characteristics, consumption risk, and firm-level risk exposures. *Journal of Financial Economics*, 125(2):326–343.

Dixon, M. (2021). Industrial forecasting with exponentially smoothed recurrent neural networks. *Technometrics*, Forhtcoming:1–27.

Dupire, B. (1994). Pricing with a smile. *Risk*, 7(1):18–20.

Eberlein, E. and Kallsen, J. (2019). *Mathematical finance*. Springer.

Engle, R. F., Ledoit, O., and Wolf, M. (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 37(2):363–375.

Falck, A., Rej, A., and Thesmar, D. (2021). Why and how systematic strategies decay. *arXiv Preprint*, (2105.01380).

Fallahgoul, H. and Loeper, G. (2021). Modelling tail risk with tempered stable distributions: an overview. *Annals of Operations Research*, 299(1):1253–1280.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance*, 47(2):427–465.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33:3–56.

Fama, E. F. and French, K. R. (2020). Comparing cross-section and time-series factor models. *Review of Financial Studies*, 33(5):1891–1926.

Fang, F. and Oosterlee, C. W. (2009). A novel pricing method for European options based on Fourier-cosine series expansions. *SIAM Journal on Scientific Computing*, 31(2):826–848.

Farmer, L., Schmidt, L., and Timmermann, A. (2021). Pockets of predictability. *SSRN Working Paper*, 3152386.

Ferreiro-Castilla, A., Kyprianou, A. E., Scheichl, R., and Suryanarayana, G. (2014). Multilevel monte carlo simulation for lévy processes based on the wiener–hopf factorisation. *Stochastic Processes and their Applications*, 124(2):985–1010.

Fournié, E., Lasry, J.-M., Lebuchoux, J., Lions, P.-L., and Touzi, N. (1999). Applications of Malliavin calculus to Monte-Carlo methods in finance. *Finance and Stochastics*, 3(4):391–412.

Frazzini, A. and Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, 111(1):1–25.

Freund, Y. and Schapire, R. E. (1996). Experiments with a new boosting algorithm. In *ICML*, volume 96, pages 148–156.

Freund, Y. and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139.

Freyberger, J., Neuhierl, A., and Weber, M. (2020). Dissecting characteristics non-parametrically. *Review of Financial Studies*, 33(5):2326–2377.

Fusai, G., Germano, G., and Marazzina, D. (2016). Spitzer identity, Wiener-Hopf factorization and pricing of discretely monitored exotic options. *European Journal of Operational Research*, 251(1):124–134.

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., and Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4):1–37.

Ge, S., Li, S., and Linton, O. (2021). Dynamic peer groups of arbitrage characteristics. *SSRN Working Paper*, 3638105.

Gehrig, T., Sögner, L., and Westerkamp, A. (2018). Making parametric portfolio policies work. *SSRN Working Paper*, 3081100.

Giglio, S., Xiu, D., and Zhang, D. (2021). Test assets and weak factors. *SSRN working Paper*, 3768081.

Giles, M. B. (2008). Multilevel monte carlo path simulation. *Operations Research*, 56(3):607–617.

Giles, M. B. and Xia, Y. (2017). Multilevel monte carlo for exponential lévy models. *Finance and Stochastics*, 21(4):995–1026.

Goetzmann, W. N. and Kumar, A. (2008). Equity portfolio diversification. *Review of Finance*, 12(3):433–463.

Gorishniy, Y., Rubachev, I., Khrulkov, V., and Babenko, A. (2021). Revisiting deep learning models for tabular data. *arXiv Preprint*, (2106.11959).

Goto, S. and Xu, Y. (2015). Improving mean variance optimization through sparse hedging restrictions. *Journal of Financial and Quantitative Analysis*, 50(6):1415–1441.

Goyal, A. and Jegadeesh, N. (2018). Cross-sectional and time-series tests of return predictability: What is the difference? *Review of Financial Studies*, 31(5):1784–1824.

Graham, B., Dodd, D. L. F., Cottle, S., et al. (1934). *Security analysis*, volume 452. McGraw-Hill New York.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5):2223–2273.

Guida, T. and Coqueret, G. (2018a). Ensemble learning applied to quant equity: gradient boosting in a multifactor framework. *Big data and machine learning in quantitative investment*, pages 129–148.

Guida, T. and Coqueret, G. (2018b). Machine learning in systematic equity allocation: A model comparison. *Wilmott*, 2018(98):24–33.

Guillaume, F. and Schoutens, W. (2012). Calibration risk: Illustrating the impact of calibration risk under the heston model. *Review of Derivatives Research*, 15(1):57–79.

Hackmann, D. and Kuznetsov, A. (2014). Asian options and meromorphic lévy processes. *Finance and Stochastics*, 18(4):825–844.

Haddad, V., Kozak, S., and Santosh, S. (2020). Factor timing. *Review of Financial Studies*, 33(5):1980–2018.

Hartzmark, S. M. and Sussman, A. B. (2019). Do investors value sustainability? A natural experiment examining ranking and fund flows. *Journal of Finance*, 74(6):2789–2837.

Harvey, C. R. (2017). Presidential address: The scientific outlook in financial economics. *The Journal of Finance*, 72(4):1399–1440.

Harvey, C. R., Liechty, J. C., Liechty, M. W., and Müller, P. (2010). Portfolio selection with higher moments. *Quantitative Finance*, 10(5):469–485.

Harvey, C. R. and Liu, Y. (2021). Uncovering the iceberg from its tip: A model of publication bias and p-hacking. *SSRN Working Paper*, 3865813.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327–343.

Hou, K., Karolyi, G. A., and Kho, B.-C. (2011). What factors drive global stock returns? *Review of Financial Studies*, 24(8):2527–2574.

Hsu, C.-C. and Chen, M.-L. (2017). The timing of low-volatility strategy. *Finance Research Letters*, 23:114–120.

Ilmanen, A. (2011). *Expected returns: An investor's guide to harvesting market rewards*. John Wiley & Sons.

Ilmanen, A., Israel, R., Moskowitz, T. J., Thapar, A. K., and Lee, R. (2021). How do factor premia vary over time? A century of evidence. *SSRN Working Paper*, 3400998.

Jackwerth, J. and Rubinstein, M. (1996). Recovering probability distributions from option prices. *Journal of Finance*, 51(5):1611–1631.

Jacod, J. and Shiryaev, A. (2013). *Limit theorems for stochastic processes*, volume 288. Springer Science & Business Media.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 58(4):1651–1683.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of finance*, 48(1):65–91.

Jensen, T. I., Kelly, B. T., and Pedersen, L. H. (2021). Is there a replication crisis in finance? *Journal of Finance*, Forthcoming.

Jondeau, E. and Rockinger, M. (2006). Optimal portfolio allocation under higher moments. *European Financial Management*, 12(1):29–55.

Jones, C. S. and Pomorski, L. (2017). Investing in disappearing anomalies. *Review of Finance*, 21(1):237–267.

Keen, S. (2020). The appallingly bad neoclassical economics of climate change. *Globalizations*, pages 1–29.

Kelly, B. T., Malamud, S., and Pedersen, L. H. (2020). Principal portfolios. *SSRN Working Paper*, 3623983.

Kelly, B. T., Pruitt, S., and Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3):501–524.

Kim, S., Korajczyk, R. A., and Neuhierl, A. (2021a). Arbitrage portfolios. *Review of Financial Studies*, 34(6):2813–2856.

Kim, S., Korajczyk, R. A., and Neuhierl, A. (2021b). Characteristic-based returns: Alpha or smart beta? *Journal of Investment Management*, Forthcoming.

Kim, Y. S., Rachev, S. T., Bianchi, M. L., and Fabozzi, F. J. (2010). Tempered stable and tempered infinitely divisible GARCH models. *Journal of Banking & Finance*, 34(9):2096–2109.

Koijen, R., Richmond, R. J., and Yogo, M. (2020). Which investors matter for equity valuations and expected returns? *SSRN Working Paper*, 3378340.

Koijen, R. S. and Yogo, M. (2019). A demand system approach to asset pricing. *Journal of Political Economy*, 127(4):1475–1515.

Kudryavtsev, O. E. and Luzhetskaya, P. (2020). The Wiener-Hopf factorization for pricing options made easy. *SSRN Working Paper*, 3540466.

Kuznetsov, A., Kyprianou, A. E., Pardo, J. C., van Schaik, K., et al. (2011). A wiener–hopf monte carlo simulation technique for lévy processes. *Annals of Applied Probability*, 21(6):2171–2190.

Kyprianou, A., Schoutens, W., and Wilmott, P. (2006). *Exotic option pricing and advanced Lévy models*. John Wiley & Sons.

Kyprianou, A. E. (2006). *Introductory lectures on fluctuations of Lévy processes with applications*. Springer Science & Business Media.

Lagerkvist, C., Edenbrandt, A., Tibbelin, I., and Wahlstedt, Y. (2020). Preferences for sustainable and responsible equity funds-a choice experiment with swedish private investors. *Journal of Behavioral and Experimental Finance*, 28:100406.

Lazar, E., Qi, S., and Tunaru, R. (2020). Measures of model risk in continuous-time finance models. *arXiv Preprint*, (2010.08113).

Leary, M. T. and Michaely, R. (2011). Determinants of dividend smoothing: Empirical evidence. *Review of Financial Studies*, 24(10):3197–3249.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.

Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics*, 40(2):1024–1060.

Ledoit, O. and Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *Review of Financial Studies*, 30(12):4349–4388.

Lettau, M. and Pelger, M. (2020a). Estimating latent asset-pricing factors. *Journal of Econometrics*, 218(1):1–31.

Lettau, M. and Pelger, M. (2020b). Factors that fit the time series and cross-section of stock returns. *Review of Financial Studies*, 33(5):2274–2325.

Maillet, B., Tokpavi, S., and Vaucher, B. (2015). Global minimum variance portfolio optimisation under some model risk: A robust regression-based approach. *European Journal of Operational Research*, 244(1):289–299.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1):77–91.

Marti, G. (2020). Corrgan: Sampling realistic financial correlation matrices using generative adversarial networks. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8459–8463.

Matache, A.-M., Von Petersdorff, T., and Schwab, C. (2004). Fast deterministic pricing of options on lévy driven assets. *ESAIM: Mathematical Modelling and Numerical Analysis*, 38(1):37–71.

McLean, R. D. and Pontiff, J. (2016). Does academic research destroy stock return predictability? *Journal of Finance*, 71(1):5–32.

Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, pages 141–183.

Michna, Z. (2011). Formula for the supremum distribution of a spectrally positive α-stable lévy process. *Statistics & Probability Letters*, 81(2):231–235.

Michna, Z. (2013). Explicit formula for the supremum distribution of a spectrally negative stable process. *Electronic Communications in Probability*, 18.

Michna, Z., Palmowski, Z., and Pistorius, M. (2015). The distribution of the supremum for spectrally asymmetric lévy processes. *Electronic Communications in Probability*, 20:1–10.

Molnar, C. (2020). *Interpretable machine learning*. Lulu. com.

Molnar, C., Casalicchio, G., and Bischl, B. (2020). Interpretable machine learning–a brief history, state-of-the-art and challenges. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 417–431. Springer.

Moreno-Torres, J. G., Raeder, T., Alaiz-Rodríguez, R., Chawla, N. V., and Herrera, F. (2012). A unifying view on dataset shift in classification. *Pattern recognition*, 45(1):521–530.

Morgan, J. N. and Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal. *Journal of the American Statistical Association*, 58(302):415–434.

Müller, K. and Schmickler, S. (2020). Interacting anomalies. *SSRN Working Paper*, 3646417.

Naranjo, A., Nimalendran, M., and Ryngaert, M. (1998). Stock returns, dividend yields, and taxes. *Journal of Finance*, 53(6):2029–2057.

Nordhaus, W. D. and Boyer, J. (2000). *Warming the world: economic models of global warming*. MIT press.

Oh, D. H. and Patton, A. J. (2017). Modeling dependence in high dimensions with factor copulas. *Journal of Business & Economic Statistics*, 35(1):139–154.

Pearl, J. (2009). *Causality*. Cambridge University Press.

Pedersen, L. H., Babu, A., and Levine, A. (2021). Enhanced portfolio optimization. *Financial Analysts Journal*, page 1.

Penasse, J. (2020). Understanding alpha decay. *SSRN Working Paper*, 2953614.

Raponi, V., Robotti, C., and Zaffaroni, P. (2020). Testing beta-pricing models using large cross-sections. *Review of Financial Studies*, 33(6):2796–2842.

Ross, L. (2020). Are characteristic interactions important to the cross-section of expected returns? *SSRN Working Paper*, 3862847.

Sato, K.-i. (1999). *Lévy processes and infinitely divisible distributions*. Cambridge University Press.

Schoutens, W. (2003). *Lévy processes in finance: Pricing financial derivatives*. Wiley.

Schwert, G. W. (2003). Anomalies and market efficiency. *Handbook of the Economics of Finance*, 1:939–974.

Shea, Y. and Radatz, E. (2020). Searching for inner peace with value factors. *Risk & Reward*, pages 14–19.

Shwartz-Ziv, R. and Armon, A. (2021). Tabular data: Deep learning is not all you need. *arXiv Preprint*, (2106.03253).

Stevens, G. V. (1998). On the inverse of the covariance matrix in portfolio analysis. *Journal of Finance*, 53(5):1821–1827.

Stroebel, J. and Wurgler, J. (2021). What do you think about climate finance? *Journal of Financial Economics*, Forthcoming.

Suh, S. (2019). A filtering strategy for improving characteristics-based portfolios. *SSRN Working Paper*, 3874684.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Tuck, J., Barratt, S., and Boyd, S. (2021). Portfolio construction using stratified models. *arXiv Preprint*, (2101.04113).

Uhl, M. W. and Rohner, P. (2018). Robo-advisors versus traditional investment advisors: An unequal game. *Journal of Wealth Management*, 21(1):44–50.

Valkanov, R. (2003). Long-horizon regressions: theoretical results and applications. *Journal of Financial Economics*, 68(2):201–232.

Windmüller, S. (2021). Firm characteristics and global stock returns: A conditional asset pricing model. *SSRN Working Paper*, 3760593.

Žliobaitė, I. (2010). Learning under concept drift: an overview. *arXiv Preprint*, (1010.4784).

Zolotarev, V. M. (1986). *One-dimensional stable distributions*, volume 65. American Mathematical Soc.