MEMOTEF

ANNALI
DEL DIPARTIMENTO
DI METODI
E MODELLI
PER L'ECONOMIA
IL TERRITORIO
E LA FINANZA

2022

University Press

SAPIENZA
UNIVERSITÀ EDITRICE

Direttrice Responsabile | ***Director*** Alessandra De Rose

Direttrice Scientifica | ***Editor in Chief*** Roberta Gemmiti

Curatrici del numero | ***Managing Editors*** Giuseppina Guagnano, Maria Rita Sebastiani

Sapienza Università Editrice
Piazzale Aldo Moro 5 – 00185 Roma

www.editricesapienza.it
editrice.sapienza@uniroma1.it

Iscrizione Registro Operatori Comunicazione n. 11420
Registry of Communication Workers registration n. 11420

Rivista di proprietà della Sapienza Università di Roma, pubblicata con il contributo dell'Ateneo
Journal owned by Sapienza Università di Roma, published with the contribution of the University

Autorizzazione del Tribunale Civile di Roma n. 105/2019 del 01/08/2019 (online) e n. 247/2016 del 30/12/2016 (print) | *Authorization of the Civil Court of Rome n.* 105/2019 del 01/08/2019 *(online) and n.* 247/2016 del 30/12/2016 *(print)*

e-ISSN: 2611-6634 | ISSN: 2385-0825

Pubblicato e stampato a dicembre 2022 | *Published and printed in December 2022*

https://rosa.uniroma1.it/annali_memotef

Impaginazione a cura di | *Layout by:* Managing Editor

SAPIENZA
UNIVERSITÀ EDITRICE

ANNALI DEL DIPARTIMENTO DI METODI E MODELLI PER L'ECONOMIA, IL TERRITORIO E LA FINANZA

2022

INDEX

Research Articles

Notes and Discussions

Book Reviews

PREFACE

We are pleased to present the 2022 Issue of the Annals of the Department of Methods and Models for Economics, Territory and Finance (MEMOTEF) of *Sapienza* University of Rome.

The Annals are a peer-review multidisciplinary journal, whose aim is to bring together scholars of different disciplines, as well as empirical studies and theoretical conceptual frameworks covering topics of international relevance and reflecting the interdisciplinary nature and advanced scientific research context of the MEMOTEF Department.

According to tradition, the Journal is structured in three sections: *Research Articles*, *Notes and Discussions* and *Book Reviews*. The first contains contributions selected following a peer review process; the second concerns papers aimed at representing a forum for thought on current research; the last offers a useful insight into new studies aimed at deepening and disseminating knowledge.

In the current Issue, the main section of *Research Articles* includes eight contributions.

The first paper, by Andrea Cinfrignini, deals with asset pricing rules and proposes a non-linear pricing rule, expressed as a Choquet expectation.

The second contribution, by Lorenzo Giammei, refers to the analysis of causal relations and suggests an integrated approach, by merging the two traditional methodologies of Potential Outcomes and Causal Graphs, and shows how the limitations of one can be compensated by the solutions provided by the other.

In the third article, Daniele Mancinelli reviews the main properties of the most used allocation algorithm, the so-called Constant Proportion Portfolio Insurance (CPPI), and describes the main extensions of CPPI proposed in the literature to improve its capability to reduce cash-in risk.

The fourth contribution, by Ilaria Stefani, reviews the literature on portfolio allocation, that is the decision-making process to determine how financial resources should be allocated among different possible investments.

The fifth paper, by Barbara Brollo, refers to temporary populations in urban contexts, in the literature usually related to tourism, study, work and lifestyle migration. The novelty of the proposal is to consider temporariness as a dimension that may characterize different populations. Through a quantitative analysis based on a specific case study of Rome, the Author also evaluates the impact of temporary populations on the city's development model.

In the sixth article, starting from the mitigation actions initiated in some glacial bodies of the Adamello-Presanella mountain range, Adriana Conti Puorger reflects on the links between sustainability and recovery. This article aims to discuss the opportunities and limitations of the digital technologies for tourism promotion and development in the declining areas of Italian mountains, adopting the methodological perspective of the so-called learning tourism.

The seventh paper, by Cinzia Di Palo and Pierluigi Fava, analyses a potential investment by a possible pension investor in government bonds issued in foreign currency under a risk-return perspective. By means of stochastic processes, the trend in the exchange rate between the domestic and foreign currency is simulated to assess its impact on the yield to maturity of the government bond considered, as well as to determine the probabilities of obtaining a negative yield or returns to maturity lower than those obtainable with a similar investment instrument.

The last paper of the section, by Giuseppe Reale, concerns the role of asymmetry in the distribution of portfolio returns and investors' preferences. The Author proposes a new different strategy that seems to be very advantageous from the theoretical point of view, as well as in terms of interpretation of role of asymmetry on investor's preferences.

In the section *Notes and Discussions* there are three papers.

In the first, Marco Brogna contributes to debate about impact of COVID-19 pandemic on tourism development model, providing some evidences concerning the dimension of tourist occupation in Italy. Since the pandemic has highlighted the many factors of structural weakness in the sector, the Author emphasizes needs for introducing significant changes in the objectives and methods of tourism development.

The second contribution, by Carla Ventre, concerns migrations and the so-called Balkan route. The Author reviews a detailed literature about the policies adopted by the European States to manage migration along Balkan countries, discussing effects of these policies in terms of changes in the route and especially on respect of the fundamental human rights.

In the last paper, Marina Zannella addresses the topic of care and well-being of elderly people, one of the most important challenges to the Italian welfare system. Noting that in Italy solidarity and silver cohousing projects are still scarce and limited to the sphere of social experimentation, the article presents some existing good practices on the territory showing their several advantages also in terms of improved inclusiveness.

The section *Book Reviews* contains a contribution by Alessandra De Rose. The book presented and discussed deals with the decreasing in fertility in Italy and the possible actions and policies to be done, so that Italy could go back to be a "country for mothers".

We thank the Authors and all who responded to our call for papers.

We are very grateful to all referees for their precious commitment in reviewing the submitted papers.

We hope that the contributions in this Issue could further stimulate interdisciplinary debate and research in the MEMOTEF Department.

December, 2022
Giuseppina Guagnano and Maria Rita Sebastiani

RESEARCH ARTICLES

SAPIENZA
UNIVERSITÀ EDITRICE

ANNALI DEL DIPARTIMENTO DI METODI E MODELLI
PER L'ECONOMIA IL TERRITORIO E LA FINANZA
2022 | pp. 11-28

E-ISSN 2611-6634 | ISSN 2385-0825
DOI: 10.13133/2611-6634/1392

Research article

Andrea Cinfrignini[*]

PRICING THROUGH THE CHOQUET INTEGRAL

Abstract

The classical no-arbitrage pricing theory allows to price assets through a linear pricing rule, by assuming a frictionless and competitive market. Moreover, completeness of the market assures that the pricing rule is defined as a discounted expected value with respect to a unique equivalent martingale measure. On the other hand, under no-arbitrage assumption, incomplete models, such as the trinomial model, lead to a set of equivalent martingale measures. This suggests to work with non-linear pricing rules that can allow frictions in the market. A generalized pricing rule can be achieved by replacing additive measures with non-additive measures such as convex capacities and belief functions in Dempster-Shafer theory. The paper recaps results on non-additive measures and Choquet expectation as non-linear functional to be used in pricing. In the literature it has been proved that, under suitable conditions, a non-linear pricing rule can be expressed as a Choquet expectation with respect to a convex capacity. In the trinomial market model the lower probability is a belief function, but it cannot be used to reach the lower expectation through the Choquet integral. Nevertheless it can avoid a generalized Dutch book condition in the framework of partially resolving uncertainty.

Keywords: *incomplete markets, non-linear pricing rule, Choquet integral, belief functions, generalized no-Dutch book.*

---

[*] Department of Methods and Models for Territory, Economics and Finance - Sapienza University of Rome

1 Introduction

The classical pricing theory is based on the assumptions that the market is frictionless and competitive. Hence, the existence of a linear pricing rule is equivalent to the fact that the market is arbitrage-free (the first fundamental theorem of asset pricing). In turn, the existence of a linear pricing rule is equivalent to the existence of an equivalent martingale measure. Moreover, the assumption of *completeness* of the market assures that the equivalent martingale measure is unique.

Under market incompleteness, the uniqueness is lost and this leads to a set of equivalent martingale measures (see, e.g., Amihud and Mendelson, 1986; Chen and Kulperger, 2006; Acciaio et al., 2016). The literature concerning the theory of sets of probability measures and their envelopes essentially refers to Walley (1991), Gilboa and Schmeidler (1989), Schmeidler (1989), Cozman (2000), Ghirardato and Marinacci (2001), Capotorti et al. (2008), Coletti et al. (2016), Erreygers et al. (2019), Petturiti and Vantaggi (2020), T'Joens et al. (2021), Petturiti and Vantaggi (2022).

In the framework of decision theory, sets of probability measures are related also to the notion of ambiguity (Etner et al., 2012; Gilboa and Marinacci, 2011).

As is well-known, the simplest example of incomplete market is the *trinomial market model*. In the classical approach the market can be completed by adding another risky asset that leads to choose a specific equivalent martingale measure in the original set. Anyhow the latter procedure requires a choice criterion and it would lead to lose some information contained in the set. More generally, incompleteness continues to hold if the risky asset is allowed to have n different possible future values, for $n \geq 3$.

The existence of a set of probability measures suggests to work with a non-linear pricing rule that can model frictions in the market. Frictions such as bid-ask spreads are largely proved to exist (Amihud and Mendelson, 1986, 1991) and they are studied in Bensaid et al. (1992), Jouini and Kallal (1995), Acciaio et al. (2016), Cerraia-Vioglio et al. (2015), Chateauneuf et al. (1996), Chateauneuf and Cornet (2022).

There are alternative attempts along this line by considering different functionals for pricing: envelopes of expected values with respect to a class of probability measures, integral forms such as Choquet expectation with respect to *non-additive measures*. In general, the two approaches are not equivalent but in case of a convex capacity (or a belief function) ν the Choquet integral coincides with the lower expectation induced by its core (see Schmeidler, 1986). In particular, in Cinfrignini et al. (2021) the study of market frictions has been faced by replacing probability measures with belief functions in the Dempster-Shafer theory (Dempster, 1967; Shafer, 1976).

The paper is structured as follows. In Section 2 we report the classical no-arbitrage pricing theory in the one-period setting. We introduce complete and incomplete markets and we show the one-period trinomial market model as a prototypical example of incomplete market. Section 3 introduces non-additive measures that are required to deal with non-linear pricing rules and the Choquet integral as non-linear functional. In Section 4 we recall and connect some results given in Chateauneuf et al. (1996) and Coletti et al. (2020) assuring that a non-linear pricing rule can be expressed through a Choquet expectation. In particular we will focus on a *global* lower pricing rule that can be expressed as a discounted Choquet expectation with respect to a convex capacity or a belief function. Then we point out that in the trinomial market model the lower probability, proved to be a belief function in Cinfrignini et al. (2021), gives rise to a Choquet expectation that does not coincide with the lower expectation induced by the equivalent martingale measures. Nevertheless, the lower price assessment on the bond and the risky asset satisfies the generalized no-Dutch-book condition obtained from Coletti et al. (2020). Finally, the last section draws conclusions.

2 Classical one-period no-arbitrage theory

We refer to a one-period financial market open at times $t = 0$ and $t = 1$. An *asset* (or security) is a tradable financial instrument that has a positive or negative cash flow of money. The cash flow is deterministic (i.e. it does not depend on future states of the world) when the asset is *riskless*; otherwise the cash flow is a random variable since it depends on what state of world will occur, and the asset is called *risky*. The market is based on two fundamental assumptions (Allingham, 1991):

(i) absence of frictions (there are no transaction costs, taxes and others restrictions on trading);

(ii) competitiveness (every quantity can be traded at market's price).

One period market model consists of a set of K risky assets with price process $(S_0^{(k)}, S_1^{(k)})$, for $k = 1, \ldots, K$, and by one riskless asset (bond) with price process (B_0, B_1) that is identified with a 0-th asset $(S_0^{(0)}, S_1^{(0)})$ to simplify the notation. It is usually assumed that $S_0^{(k)} = s^{(k)} > 0$ is a deterministic positive value (called *price*), while $S_1^{(k)}$ is a random variable (called *payoff*), for each $k = 1, \ldots, K$. The bond process, without loss of generality, is assumed to be $S_0^{(0)} = 1$ and $S_1^{(0)} = 1 + r$, where $r > 0$ is the risk-free interest rate of the market.

Price processes are defined on a filtered probability space $(\Omega, \{\mathcal{F}_0, \mathcal{F}_1\}, \mathcal{F}, P)$ where $\Omega = \{1, \ldots, n\}$, $n \in \mathbb{N}$ is a finite state space, $\{\mathcal{F}_0, \mathcal{F}_1\}$ is a filtration such that $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_1 = \mathcal{F} = \mathcal{P}(\Omega)$ is the power set of Ω, and P is a probability measure on $\mathcal{F}$. The probability measure is called *"natural"* or *"real-world"* probability measure and the classical pricing theory asks for the positivity of P since it assures that an asset with a non-negative and non-null payoff will have a positive price at time $t = 0$. We also denote by $\mathbb{R}^\Omega$ the set of all random variables which are automatically $\mathcal{F}$-measurable. Moreover, scalar real numbers are identified with constant random variables. Finally, $\mathbf{P}(\Omega, \mathcal{F})$ stands for the set of all probability measures on $(\Omega, \mathcal{F})$.

Let us denote the set of all random payoff with $\mathcal{G} = \{S_1^{(0)}, \ldots, S_1^{(K)}\}$ and with $\pi : \mathcal{G} \to \mathbb{R}$ a function such that $\pi\left(S_1^{(k)}\right) = S_0^{(k)}$, for $k = 0, \ldots, K$, which is called *price assessment*. Our aim is to look for a global *pricing rule* $\pi' : \mathbb{R}^\Omega \to \mathbb{R}$ that extends π.

The risk-free bond is usually used as a numraire (see Pliska, 1997); it means that the riskless bond allows to discount the risky process and defines a new process denoted as $\left(\tilde{S}_0^{(k)}, \tilde{S}_1^{(k)}\right)$ with $\tilde{S}_0^{(k)} = S_0^{(k)}$ and $\tilde{S}_1^{(k)} = (1 + r)^{-1} S_1^{(k)}$, for $k = 1, \ldots, K$.

A *portfolio* (or *trading strategy*) is a collection of assets that an agent can hold. It is denoted by a vector $\boldsymbol{\lambda} = (\lambda_0, \ldots, \lambda_K) \in \mathbb{R}^{K+1}$, whose component λ_k expresses the number of units purchased

$(\lambda_k > 0)$ or sold $(\lambda_k < 0)$ of the k-th asset in the time interval $[0, 1]$.

The *price* at time $t = 0$ of the portfolio $\boldsymbol{\lambda}$ is computed as weighted sum of prices:

$$V_0^{\boldsymbol{\lambda}} = \sum_{k=0}^{K} \lambda_k S_0^{(k)} = \sum_{k=0}^{K} \lambda_k \pi \left(S_1^{(k)} \right); \tag{1}$$

while the *payoff* of the portfolio $\boldsymbol{\lambda}$ is given by a random variable $V_1^{\boldsymbol{\lambda}} : \Omega \to \mathbb{R}$ defined, for every $i \in \Omega$, as the weighted sum of payoffs:

$$V_1^{\boldsymbol{\lambda}}(i) = \sum_{k=0}^{K} \lambda_k S_1^{(k)}(i). \tag{2}$$

Given the set of random variables $\mathcal{G}$, $\boldsymbol{\lambda} \in \mathbb{R}^{K+1}$ is a *Dutch-book* portfolio if the following condition holds:

$$\max_{i \in \Omega} \sum_{k=0}^{K} \lambda_k \left(\tilde{S}_1^{(k)}(i) - \pi \left(S_1^{(k)} \right) \right) < 0. \tag{3}$$

The condition means that the portfolio $\boldsymbol{\lambda} \in \mathbb{R}^{K+1}$ gives rise to a *sure loss* for each $i \in \Omega$, since the supremum gain is negative for sure. The portfolio is also called *incoherent*. Conversely, if inequality in Equation (3) does not hold, the portfolio is called *coherent* and it avoids a *Dutch-book* opportunity, i.e. it avoids a sure loss (Schervish et al., 2008).

The arbitrage definition is stronger than that of Dutch-book, since the former guarantees a positive payoff in, at least, one state of the world, with a zero or negative price. A portfolio $\boldsymbol{\lambda} \in \mathbb{R}^{K+1}$ is an arbitrage portfolio if one of the following condition holds (Allingham, 1991):

(1) $V_0^{\boldsymbol{\lambda}} \leq 0$ and $V_1^{\boldsymbol{\lambda}} \geq 0$ with a strict inequality for at least one $i \in \Omega$;

(2) $V_0^{\boldsymbol{\lambda}} < 0$ and $V_1^{\boldsymbol{\lambda}} = 0$.

Equivalently $\boldsymbol{\lambda} \in \mathbb{R}^{K+1}$ is an arbitrage portfolio if $\sum_{k=0}^{K} \lambda_k \left(\tilde{S}_1^{(k)}(i) - \pi \left(S_1^{(k)} \right) \right) \geq 0$, for all i, with a strict inequality for at least one $i \in \Omega$. Note that a Dutch-book opportunity implies the existence of an arbitrage but the converse does not hold (Schervish et al., 2008).

The assumption that the market has to be *arbitrage-free* is standard in classical pricing theory (see, e.g., Pliska, 1997; Dybvig and Ross, 1989) and it has important implications in asset pricing. The absence of arbitrage opportunities guarantees the existence of a positive linear pricing rule $\pi' : \mathbb{R}^{\Omega} \to \mathbb{R}$ such that $\pi' \left(S_1^{(k)} \right) = \pi \left(S_1^{(k)} \right)$, for $k = 0, \ldots, K$ (Dybvig and Ross, 1989).

Furthermore, when the market is *complete*, there is a unique linear pricing rule π' given by the discounted expected value computed with respect to a unique risk-neutral probability measure that has to be equivalent to the natural one.

Under completeness, a *derivative* X, that is a financial contract defined as a random process (X_0, X_1) on the filtered probability space $(\Omega, \{\mathcal{F}_0, \mathcal{F}_1\}, \mathcal{F}, P)$, adapted to the filtration $\{\mathcal{F}_0, \mathcal{F}_1\}$, can be perfectly replicated by setting up a replicating strategy $\boldsymbol{\lambda} \in \mathbb{R}^{K+1}$ composed by the risky assets and the bond, such that they have the same final payoff $X_1 = V_1^{\boldsymbol{\lambda}}$.

Then, by the *law of one price*, they have the same price at time $t = 0$:

$$X_0 = V_0^{\boldsymbol{\lambda}}, \tag{4}$$

and its value is computed as discounted expected value of its payoff:

$$X_0 = (1+r)^{-1}\mathbb{E}_Q(X_1), \tag{5}$$

where Q is the unique equivalent martingale measure. Therefore, we have that $\pi'(\cdot) = (1+r)^{-1}\mathbb{E}_Q(\cdot)$.

On the other hand, in the case of an incomplete market, the price assessment is consistent with the no-arbitrage assumption but not each derivative in the market can be replicated by a strategy. This leads to a set of equivalent martingale measures $\mathcal{Q}$ such that each $Q \in \mathcal{Q}$ defines a different price.

Given a *non-replicable* derivative with payoff $Y_1 \in \mathbb{R}^{\Omega}$, its fair price can be computed as an interval defined through the closest replicable derivative. If X_1 is the closest replicable derivative of Y_1, the following quantities can be computed:

$$\overline{V}(Y_1) = \inf_{\substack{X_1 \le Y_1, \\ X_1 \text{ is replicable}}} (1+r)^{-1}\mathbb{E}_Q(X_1), \quad \underline{V}(Y_1) = \sup_{\substack{X_1 \le Y_1, \\ X_1 \text{ is replicable}}} (1+r)^{-1}\mathbb{E}_Q(X_1). \tag{6}$$

The fair price of the derivative has to be in the interval $\big(\underline{V}(Y_1), \overline{V}(Y_1)\big)$, otherwise it gives rise to an arbitrage opportunity (Pliska, 1997). Another approach to select a replicating strategy for a non-replicable derivative is to choose the best replicating strategy among the imperfect strategies through approximations/algorithms (see Cern, 2009; Bertsimas et al., 2001). Although they are not detailed here, some criteria to choose a replicating strategy can be the following:

(a) ***sub(super)-hedging.*** We look for a strategy $\boldsymbol{\lambda}_S \in \mathbb{R}^{K+1}$ such that $V_1^{\boldsymbol{\lambda}_S} \le (\ge) Y_1$. Hence the sub-hedging $\underline{V}_0^{\boldsymbol{\lambda}_S}$ and the super-hedging prices $\overline{V}_0^{\boldsymbol{\lambda}_S}$ are the no-arbitrage bounds for the non-replicable payoff Y_1;

(b) ***quadratic risk minimization.*** We look for a strategy $\boldsymbol{\lambda}_{QR} \in \mathbb{R}^{K+1}$ that minimizes the expected value of the quadratic distance between the payoff of the derivative and the value of the portfolio. The following optimization problem has to be solved:

$$\min_{\boldsymbol{\lambda}_{QR}} \mathbb{E}\left[\left(Y_1 - V_1^{\boldsymbol{\lambda}_{QR}}\right)^2\right]; \tag{7}$$

(c) ***shortfall risk minimization.*** We look for a strategy $\boldsymbol{\lambda}_{SR} \in \mathbb{R}^{K+1}$ that minimizes the shortfall risk. It penalizes only deviations in defect but it is less mathematically tractable. The following problem has to be solved:

$$\min_{\boldsymbol{\lambda}_{SR}} \mathbb{E}\left[\left(Y_1 - V_1^{\boldsymbol{\lambda}_{SR}}\right)^+\right]. \tag{8}$$

Another approach to overcome market's incompleteness is to *complete* the market with an appropriate number of extra assets. Let us introduce the matrix notation to go deep into the problem. Payoffs of riskless and risky assets are defined in the matrix $A \in \mathbb{R}^{n\times(K+1)}$:

$$A = \begin{bmatrix} S_1^{(0)}(1) & S_1^{(1)}(1) & \dots & S_1^{(K)}(1) \\ \vdots & \vdots & & \vdots \\ S_1^{(0)}(n) & S_1^{(1)}(n) & \dots & S_1^{(K)}(n) \end{bmatrix}, \tag{9}$$

and the vector of payoff of the derivative is denoted by $\boldsymbol{X} = (X_1(1), \dots, X_1(n)) \in \mathbb{R}^n$. Hence, an arbitrage-free market is complete if and only if the following linear problem has a unique solution:

$$A\boldsymbol{\lambda}^T = \boldsymbol{X}, \tag{10}$$

where $\boldsymbol{\lambda} = (\lambda_0, \dots, \lambda_k) \in \mathbb{R}^{K+1}$ is the portfolio such that λ_0 is referred to units of risk-free asset $S^{(0)}$ and λ_k is referred to units of risky asset $S^{(k)}$, for $k = 1, \dots, K$. Problem (10) has a unique solution, assumed that there are no *redundant* assets[1], if and only if $rank(A) = n = K + 1$, hence A has to be a square matrix. Otherwise, the following possibilities can occur (Cern, 2009):

(I1) $rank(A) = n < (K + 1)$: the market is *complete* but there are $K + 1 - n$ redundant assets that lead to $K + 1 - n$ free parameters referred to redundant assets;

(I2) $rank(A) = (K + 1) < n$: the market is *incomplete* since $n - (K + 1)$ assets are lacking. It can be completed by adding the missing number of assets;

(I3) $rank(A) < n, rank(A) < (K+1)$: the market is *incomplete* and there are $(K+1)-rank(A)$ redundant assets.

However, sometimes the completion is not possible or not desirable as it changes the market structure. Also other procedures that introduce additional requirements such as agents' preferences may be not desirable as they change the framework. A way to define a unique price consistent with the no-arbitrage principle without changing the market structure is to compute prices with every $Q \in \mathcal{Q}$ through Equation (5) and define a set of prices $\mathcal{Y}$. Then we could choose the price $Y \in \mathcal{Y}$ that departs as little as possible from the actually observed in the market. For this procedure we can refer, for example, to Pascucci and Runggaldier (2011).

The simplest example of complete market model is the one-period binomial model (Cox et al., 1979), while an example of incomplete market model is the trinomial model.

The trinomial model is composed by a bond with price process
$(B_0 = 1, B_1 = (1 + r)B_0)$ and by a risky asset with the following price process:

$$S_0 = s > 0, \quad S_1 = \begin{cases} uS_0 & \text{with probability } p_1, \\ mS_0 & \text{with probability } p_2, \\ dS_0 & \text{with probability } p_3, \end{cases} \tag{11}$$

---

[1] An asset whose payoff can be written as a linear combination of others assets' payoffs is called redundant since it does not add anything new to the market. If there are no redundant assets, the market's asset are said to be *linearly independent*.

where $u > m > d > 0$ are parameters, $p_i \in (0,1)$ for $i = 1,2,3$, and $\sum_{i=1}^{3} p_i = 1$. Such model is free of arbitrage if and only if $u > (1+r) > d$ as the binomial model but it is not complete since it occurs condition (I2).

In the trinomial case there is a set of equivalent martingale measures denoted as:

$$\mathcal{Q} = \{Q \in \mathbf{P}(\Omega, \mathcal{F}) : (1+r)^{-1}\mathbb{E}_Q(S_1) = S_0, \;\; Q \sim P\}. \tag{12}$$

The set $\mathcal{Q}$ is a convex set that can be characterized by its extreme points (Runggaldier, 2006) (in particular it is a segment since there are two extreme points):

$$Q^1 = (q_1^1, q_2^1, q_3^1) \;\; = \;\; \begin{cases} \left(0, \frac{(1+r)-d}{m-d}, \frac{m-(1+r)}{m-d}\right) & \text{if } m \geq (1+r), \\ \left(\frac{(1+r)-m}{u-m}, \frac{u-(1+r)}{u-m}, 0\right) & \text{if } m < (1+r), \end{cases} \tag{13}$$

$$Q^2 = (q_1^2, q_2^2, q_3^2) \;\; = \;\; \left(\frac{(1+r)-d}{u-d}, 0, \frac{u-(1+r)}{u-d}\right). \tag{14}$$

We stress that extreme points Q^1 and Q^2 are not equivalent to P since they are not positive on $\mathcal{F}$; hence equivalent martingale measures are given by the strict convex combinations of Q^1 and Q^2:

$$\mathcal{Q} = \{Q^\alpha : Q^\alpha = \alpha Q^1 + (1-\alpha)Q^2, \;\; \alpha \in (0,1)\}, \tag{15}$$

with $Q^\alpha \sim P$, for each $Q^\alpha \in \mathcal{Q}$.

At this point, a suitable criterion to choose one measure in the set is required. In the following example we show that each $Q^\alpha \in \mathcal{Q}$ is an equivalent martingale measure consistent with no-arbitrage assumption but it leads to varied prices for the derivative, through Equation (5).

Example 2.1 *Let $S_0 = 100, u = 2, m = \frac{6}{5}, d = \frac{2}{5}$ and, without loss of generality, $r = 0$. Extreme points of the set $\mathcal{Q}$, computed with Equations* (13) (14)*, are:*

$$Q^1 = \left(0, \frac{3}{4}, \frac{1}{4}\right), \quad Q^2 = \left(\frac{3}{8}, 0, \frac{5}{8}\right).$$

Then the set of equivalent martingale measures is given by:

$$\mathcal{Q} = \left\{Q^\alpha : Q^\alpha = \alpha\left(0, \frac{3}{4}, \frac{1}{4}\right) + (1-\alpha)\left(\frac{3}{8}, 0, \frac{5}{8}\right), \;\; \alpha \in (0,1)\right\}.$$

For instance, let $\alpha = 0.2$. The equivalent martingale measure is $Q^{0.2} = \left(\frac{6}{20}, \frac{3}{20}, \frac{11}{20}\right)$ and we can verify that $Q^{0.2} \in \mathcal{Q}$ by computing the following expected value:

$$\mathbb{E}_{Q^{0.2}}\left(\frac{S_1}{S_0}\right) = 2 \cdot \frac{6}{20} + \frac{6}{5} \cdot \frac{3}{20} + \frac{2}{5} \cdot \frac{11}{20} = 1.$$

Let C be a European call option with payoff $C_1 = \max(S_1 - K, 0)$ and strike price $K = 110$. The payoff at time $t = 1$ is the following:

$$C_1(i) = \begin{cases} 90 & \textit{if } i = 1 \\ 10 & \textit{if } i = 2 \\ 0 & \textit{if } i = 3 \end{cases}$$

The price of the call option C_0 computed through $Q^{0.2}$ is:

$$C_0 = \mathbb{E}_{Q^{0.2}}(C_1) = 90 \cdot \frac{6}{20} + 10 \cdot \frac{3}{20} = \frac{57}{2} = 28.5.$$

Let be $\alpha = 0.9$. The equivalent martingale measure is $Q^{0.9} = \left(\frac{3}{80}, \frac{27}{40}, \frac{23}{80}\right)$. Also in this case $Q^{0.9} \in \mathcal{Q}$ since $\mathbb{E}_{Q^{0.9}}\left(\frac{S_1}{S_0}\right) = 1$, and the price of the call option computed trough $Q^{0.9}$ is:

$$C_0 = \mathbb{E}_{Q^{0.9}}(C_1) = 90 \cdot \frac{3}{80} + 10 \cdot \frac{27}{40} = \frac{81}{8} = 10.125.$$

♦

The trinomial model can be completed by adding another risky asset. We denote risky assets as $S^{(1)}$ and $S^{(2)}$, each of them with price process as in (11), with parameters u_i, m_i, d_i, for $i = 1, 2$. The model is complete as $K + 1 = 3 = n$, with a unique solution for q_1, q_2, q_3 (for details see Pascucci and Runggaldier, 2011).

We stress that any n-nomial market model composed by K risky asset is incomplete, for $n \geq 3$ and $K < (n - 1)$, as explained in Cinfrignini et al. (2021).

Anyhow completing the market is not always possible or desirable. Our approach would deal with a subset $\mathcal{Q}' \subseteq \mathcal{Q}$, possibly with an equality. Pricing with $\mathcal{Q}'$ would allow to model frictions in the market in the form of bid-ask spreads. The intuitive way to face the problem of frictions in a trinomial model is to define the interval of derivative's price induced by $\mathcal{Q}'$. It means that we look for the lower and the upper bounds of price, defined as:

$$\underline{X_0} = (1+r)^{-1} \inf_{Q^\alpha \in \mathcal{Q}'} \mathbb{E}_{Q^\alpha}(X_1), \quad \overline{X_0} = (1+r)^{-1} \sup_{Q^\alpha \in \mathcal{Q}'} \mathbb{E}_{Q^\alpha}(X_1). \tag{16}$$

Thus, we could look for a lower/upper pricing rule which is given by the lower/upper envelope of a class of expectations with respect to each $Q^\alpha \in \mathcal{Q}'$ and extends the fixed lower/upper price assessment.

3 Non-additive measures and non-linear functionals

When uncertainty is not quantifiable in a single probability measure and we have to deal with a set of them, we are facing a situation called *ambiguity*. Since working with the whole class of probabilities is hard, we usually consider the envelopes of the class. For instance, in the trinomial model just defined, we would consider a lower pricing rule expressed by a functional of an envelope of the set of equivalent martingale measures. In particular, in what follows, we work with the *lower* envelope, but we point out that the upper envelope leads to the same results, since they are conjugate functions. Generally, envelopes of a set of probability measures are no longer probabilities. Hence, we have to introduce generalized functions that lose the additive property: for that they are called non-additive measures. Moreover, in particular settings, there exists a link between the envelopes of linear functionals defined with respect to a class of probability measures and a non-linear functionals computed with respect to a non-additive measure, as we show in this section.

Let $(\Omega, \mathcal{F})$ be the finite space defined in the previous section, with $\mathcal{F} = \mathcal{P}(\Omega)$.

Definition 3.1 A function $\nu : \mathcal{F} \to \mathbb{R}$ is called a *non-additive measure* or a *capacity* if it is:

(i) *normalized*: $\nu(\emptyset) = 0$ and $\nu(\Omega) = 1$;

(ii) *monotone*: $\nu(A) \leq \nu(B)$ for all $A, B \in \mathcal{F}$, with $A \subseteq B$.

Moreover, a capacity ν is called:

(a) *2-monotone* or *convex* capacity if, for every $A, B \in \mathcal{F}$:

$$\nu(A \cup B) \geq \nu(A) + \nu(B) - \nu(A \cap B); \tag{17}$$

(b) *totally monotone* capacity or *belief function* (usually denoted by Bel) if, for every $A_1, \ldots, A_k \in \mathcal{F}$ with $k \geq 2$, it holds that:

$$\nu\left(\bigcup_{i=1}^{k} A_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} \nu\left(\bigcap_{i \in I} A_i\right); \tag{18}$$

(c) (coherent) *lower probability* if there exists a set $\mathcal{P}$ of probability measures on $\mathcal{F}$ such that, for every $A \in \mathcal{F}$:

$$\nu(A) = \inf_{P \in \mathcal{P}} P(A); \tag{19}$$

(d) *probability measure* if $\nu(A \cup B) = \nu(A) + \nu(B)$, for every disjoint $A, B \in \mathcal{F}$.

If ν is a belief function, then it is also a 2-monotone capacity and a (coherent) lower probability. In turn, if ν is a probability measure, then it is also a belief function. Conversely, the property of being a lower probability does not imply 2-monotonicity and, so, neither total monotonicity.

We denote by $\mathbf{V}(\Omega, \mathcal{F})$ and $\mathbf{B}(\Omega, \mathcal{F})$, respectively, the set of all capacities and that of all belief functions on $(\Omega, \mathcal{F})$, and we stress that $\mathbf{P}(\Omega, \mathcal{F}) \subseteq \mathbf{B}(\Omega, \mathcal{F}) \subseteq \mathbf{V}(\Omega, \mathcal{F})$[2].

---

[2] Every capacity ν has a conjugate function called *dual capacity*. In general it is defined as $\overline{\nu}(A) = 1 - \nu(A^C), \forall A \in \mathcal{F}$. The dual of a lower probability is said *upper probability*; the dual of a 2-monotone (convex) capacity is said 2-*alternating (concave)* capacity; the dual of a belief function is said *plausibility* function (Pl); the dual of a probability is itself.

For every 2-monotone capacity there exists a set of dominating probability measures called *core* (or credal set) (Gilboa and Schmeidler, 1994; Walley, 1991):

$$\mathbf{core}(\nu) = \{P \in \mathbf{P}(\Omega, \mathcal{F}) | P(A) \geq \nu(A),\ \forall A \in \mathcal{F}\}. \tag{20}$$

A coherent lower probability $\underline{P}$ is such that $\mathbf{core}(\underline{P}) \neq \emptyset$ and $\underline{P}$ is its lower envelope: $\underline{P}(A) = \min_{P \in \mathbf{core}(\underline{P})} P(A)$, $\forall A \in \mathcal{F}$. In turn, a belief function, as it is a particular lower probability, can be regarded as the lower envelope of its core:

$$Bel(A) = \min_{P \in \mathbf{core}(Bel)} P(A). \tag{21}$$

Every capacity ν can be characterized in terms of another function called *Möbius inverse* (Chateauneuf and Jaffray, 1989):

$$m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \nu(B), \qquad \nu(A) = \sum_{B \subseteq A} m(B). \tag{22}$$

Proposition 3.1 (Chateauneuf and Jaffray, 1989). Given a function $\nu : \mathcal{F} \to \mathbb{R}$, let m be its Möbius inverse. Then:

(a) ν is a capacity if and only if:

$m(\emptyset) = 0$,

$\sum_{B \in \mathcal{F}} m(B) = 1$, and

$\sum_{\{i\} \in B \subseteq A} m(B) \geq 0$, for all $A \in \mathcal{F}$ and for all $i \in A$;

(b) ν is a 2-monotone capacity if and only if condition (a) holds and $\forall A \in \mathcal{F}$, and $\{i, j\} \in A$ with $i \neq j$, $\sum_{\{i,j\} \subseteq B \subseteq A} m(B) \geq 0$;

(c) ν is a belief function if and only if condition (a) holds and m is non-negative;

(d) ν is a probability measure if and only if condition (a) holds, m is non-negative and can be positive only on singletons.

Definition 3.2 (Gilboa and Schmeidler, 1994). Given a capacity ν and a random variable $X \in \mathbb{R}^{\Omega}$, the *Choquet expectation* of X with respect to ν, denoted by $\mathbb{C}_{\nu}(X)$, is defined through the Choquet integral:

$$\begin{aligned} \mathbb{C}_{\nu}(X) &= \oint_{\Omega} X \,\mathrm{d}\nu = \\ &= \int_0^{\infty} \nu(\{i \in \Omega | X(i) \geq x\}) \,\mathrm{d}x + \int_{-\infty}^{0} [\nu(\{i \in \Omega | X(i) \geq x\}) - \nu(\Omega)] \,\mathrm{d}x. \end{aligned} \tag{23}$$

We point out that the Choquet expectation coincides with the expected value if ν is additive (i.e. it is a probability measure P): $\mathbb{C}_\nu(X) = \mathbb{E}_P(X)$. Assuming $\Omega = \{1, \dots, n\}$, the Choquet integral can be computed in the following way:

$$\mathbb{C}_\nu(X) = \sum_{i=1}^{n} [X(\sigma(i)) - X(\sigma(i+1))]\, \nu(E_i^\sigma), \tag{24}$$

where σ is a permutation of Ω such that $X(\sigma(1)) \geq \ldots \geq X(\sigma(n))$, $E_i^\sigma = \{\sigma(1), \dots, \sigma(i)\}$, for $i = 1, \dots, n$, and $X(\sigma(n+1)) = 0$. Moreover, for every $\nu \in \mathbf{V}(\Omega, \mathcal{F})$ with corresponding Möbius inverse m, and $X \in \mathbb{R}^\Omega$, the Choquet expectation of X with respect to ν can be computed through the Möbius inverse:

$$\mathbb{C}_\nu(X) = \sum_{B \in \mathcal{F} \setminus \{\emptyset\}} m(B) \min_{i \in B} X(i). \tag{25}$$

We summarize some properties of the Choquet integral:

(i) for all $A \subseteq \Omega$ we have that $\mathbb{C}_\nu(\mathbf{1}_A) = \nu(A)$, with $\mathbf{1}_A : \Omega \to \{0, 1\}$ the indicator function of A such that $\mathbf{1}_A(i) = 1$ if $i \in A$ and $\mathbf{1}_A = 0$ otherwise;

(ii) for any capacities $\nu, \varphi, \in \mathbf{V}(\Omega, \mathcal{F})$ and $\alpha, \beta \in \mathbb{R}$, it holds that $\mathbb{C}_{\alpha\nu+\beta\varphi}(X) = \alpha\mathbb{C}_\nu(X) + \beta\mathbb{C}_\varphi(X)$;

(iii) (non-negative homogeneity) for any capacity ν and all $\alpha \geq 0$, it holds that $\mathbb{C}_\nu(\alpha X) = \alpha\mathbb{C}_\nu(X)$;

(iv) (constant additivity) for any capacity ν and all $\alpha \in \mathbb{R}$, it holds that $\mathbb{C}_\nu(\alpha + X) = \alpha + \mathbb{C}_\nu(X)$;

(v) (monotonicity) for any capacity ν and for any $X, Y \in \mathbb{R}^\Omega$ such that $X \leq Y$, it holds that $\mathbb{C}_\nu(X) \leq \mathbb{C}_\nu(Y)$;

(vi) if ν is a 2-monotone capacity, for any $X, Y \in \mathbb{R}^\Omega$, the Choquet integral is super-additive: $\mathbb{C}_\nu(X + Y) \geq \mathbb{C}_\nu(X) + \mathbb{C}_\nu(Y)$, and, for every $X \in \mathbb{R}^\Omega$, the Choquet expectation equals the lower expectation with respect to the $\mathbf{core}(\nu)$ (Gilboa and Schmeidler, 1994):

$$\mathbb{C}_\nu(X) = \min_{P \in \mathbf{core}(\nu)} \sum_{i \in \Omega} P(\{i\}) X(i) = \min_{P \in \mathbf{core}(\nu)} \mathbb{E}_P(X); \tag{26}$$

(vii) if ν reduces to a belief function Bel, for any $X_1, \dots, X_k \in \mathbb{R}^\Omega$, the Choquet integral is completely monotone:

$$\mathbb{C}_{Bel}\left(\bigvee_{i=1}^{k} X_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1,\dots,k\}} (-1)^{|I|+1} \mathbb{C}_{Bel}\left(\bigwedge_{i \in I} X_i\right). \tag{27}$$

We stress that property (vi) continues to hold if ν reduces to a belief function and it can be interpreted as a *lower expectation*. This shows that Choquet expectation with respect to a 2-monotone capacity or a belief function leads to a specific functional inside the class of envelopes of expectations.

In the following example we show that, despite the lower envelope of a set $\mathcal{P}$ of probability measures is 2-monotone (or even a belief function), the corresponding Choquet expectation may not coincide with the lower expectation induced by $\mathcal{P}$ if its closed convex hull does not coincide with $\mathbf{core}(\underline{P})$.

Example 3.1 *Let $\Omega = \{1, 2, 3\}$ and X be a random variable that assumes the following values: $X(i) = i$, for $i = 1, 2, 3$. Let $\mathcal{P}$ be a set of three probability measures: $\mathcal{P} = \{P_1, P_2, P_3\}$ taking values reported below:*

$\mathcal{F}$	$\emptyset$	1	2	3	12	13	23	Ω
P_1	0	1/2	1/4	1/4	3/4	3/4	1/2	1
P_2	0	1/3	1/3	1/3	2/3	2/3	2/3	1
P_3	0	2/5	2/5	1/5	4/5	3/5	3/5	1

The lower probability $\underline{P}(A) = \min_{P \in \{P_1,P_2,P_3\}} P(A), \forall A \in \mathcal{F}$, and its Möbius inverse m are reported in the following table:

$\mathcal{F}$	$\emptyset$	1	2	3	12	13	23	Ω
P_1	0	1/2	1/4	1/4	3/4	3/4	1/2	1
P_2	0	1/3	1/3	1/3	2/3	2/3	2/3	1
P_3	0	2/5	2/5	1/5	4/5	3/5	3/5	1
$\underline{P}$	0	1/3	1/4	1/5	2/3	3/5	1/2	1
m	0	1/3	1/4	1/5	1/12	1/15	1/20	1/60

Since $m(A) \geq 0$ for every $A \in \mathcal{F}$, the lower probability $\underline{P}$ is a belief function.

The Choquet integral of X with respect to $\underline{P}$, generally, is not equal to the lower expectation of X computed among $P \in \mathcal{P}$ since the convex hull $\mathbf{conv}(\mathcal{P})$, which is closed as $\mathcal{P}$ is finite, does not coincide with $\mathbf{core}(\underline{P})$. To show that, we compute the extreme points of $\mathbf{core}(\underline{P})$. Extreme points of $\mathbf{core}(\underline{P})$ are computed in the following way: for any permutation of indices $\sigma = (\sigma(1), \ldots, \sigma(n))$, extreme points are computed as $P^\sigma = (P(\sigma(1)), \ldots, P(\sigma(n)))$ with $P(\sigma(i)) = \underline{P}(\{\sigma(1), \ldots, \sigma(i)\}) - \underline{P}(\{\sigma(1), \ldots, \sigma(i-1)\})$.

Hence, for any permutation of $\{1, 2, 3\}$, we have the following set of extreme points $\mathrm{ext}(\mathbf{core}(\underline{P}))$:

$$P^{(1,2,3)} = \left(\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}\right) = P_2, \qquad P^{(1,3,2)} = \left(\tfrac{1}{3}, \tfrac{2}{5}, \tfrac{4}{15}\right),$$

$$P^{(2,1,3)} = \left(\tfrac{5}{12}, \tfrac{1}{4}, \tfrac{1}{3}\right), \qquad P^{(2,3,1)} = \left(\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{4}\right) = P_1,$$

$$P^{(3,1,2)} = \left(\tfrac{2}{5}, \tfrac{2}{5}, \tfrac{1}{5}\right) = P_3, \qquad P^{(3,2,1)} = \left(\tfrac{1}{2}, \tfrac{3}{10}, \tfrac{1}{5}\right).$$

This proves that $\mathbf{conv}(\mathcal{P}) \neq \mathbf{core}(\underline{P})$*, hence the lower expectation with respect to* $\mathcal{P}$ *and the Choquet integral with respect to* $\underline{P}$*, generally, lead to different results:*

$$\underline{\mathbb{E}}(X) = \min_{P \in \{P_1, P_2, P_3\}} \mathbb{E}_P(X) = \min\{1.75, 2, 1.8\} = 1.75,$$

$$\mathbb{C}_{\underline{P}}(X) = (3-2)\underline{P}(3) + (2-1)\,\underline{P}(23) + (1-0)\underline{P}(\Omega) = 1.7.$$

Therefore, we get that $\mathbb{C}_{\underline{P}}(X) < \underline{\mathbb{E}}(X)$*, since* $\mathbf{conv}(\mathcal{P}) \subset \mathbf{core}(\underline{P})$*.* ◆

At this point the question that arises is if a n-nomial model leads to analogous results. This problem has been faced in Cinfrignini et al. (2021). It is proved that any n-nomial market model, for $n \geq 3$ is incomplete and the lower envelope of the set of equivalent martingale measures is a belief function but the closure of the set of equivalent martingale measures does not coincide with the core of its lower envelope.

4 Non-linear pricing rules

Let us consider a one-period financial market with frictions in the form of bid-ask spreads, that can be due to the presence of intermediaries, taxes, or to the incompleteness of the market. The market consists of a risk-free bond B and of a set of K risky assets with payoffs $S_1^{(1)}, \ldots, S_1^{(K)}$.

For $k = 1, \ldots, K$, each asset's price is defined through an interval $[\underline{S}_0^{(k)}, \overline{S}_0^{(k)}]$, where $\underline{S}_0^{(k)}$ is called *bid price* and $\overline{S}_0^{(k)}$ is called *ask price* (it is tacit that $\underline{S}_0^{(k)} \leq \overline{S}_0^{(k)}$ where equality holds only if the k-th asset is frictionless). The bond price process is $(B_0 = 1, B_1 = 1 + r)$ and it is frictionless, i.e. $\underline{B}_0 = \overline{B}_0 = B_0$.

The problem is to determine non-linear functionals able to characterize bid and ask prices. In Acciaio et al. (2016), for instance, lower and upper expectations are used as non-linear functionals. The question that in literature has been addressed is if non-linear functionals can be defined by means of the lower/upper expectation with respect to a set of probabilities, or by means of a Choquet integral with respect to a 2-monotone capacity or a belief function, and if the two approaches give out to the same outcome.

In what follows we will see that the same question arises in the trinomial model where a generalization of no-Dutch book condition can be shown to hold. We point out that the framework in Chateauneuf et al. (1996) and Coletti et al. (2020) is from the upper price point of view, in terms of concave capacities and plausibility functions. Here, the setting has been reversed in terms of convex capacities and belief functions. Original results do not change since concave (plausibility) functions are the conjugate of convex (belief) functions.

We consider a *global* lower pricing rule $\underline{\pi} : \mathbb{R}^\Omega \to \mathbb{R}$ defined for all $X_1 \in \mathbb{R}^\Omega$ as:

$$\underline{\pi}(X_1) = \underline{X}_0, \tag{28}$$

which is not assumed to be linear.

In accordance with Chateauneuf et al. (1996)[3], we make the following assumptions:

---

[3] In the quoted paper, the authors consider already discounted amounts that, in our setting, is equivalent to take $r = 0$.

(A1) *monotonicity*: for $X_1, Y_1 \in \mathbb{R}^{\Omega}$, if $X_1 \geq Y_1$ then $\underline{\pi}\left(X_1\right) \geq \underline{\pi}\left(Y_1\right)$;

(A2) *frictionless bond*: there is a risk-free bond $B_0 = 1, B_1 = 1 + r$ that is not frictional $\underline{\pi}(\alpha B_1) = \alpha$, for all $\alpha \in \mathbb{R}$;

(A3) *super-additivity*: for $X_1, Y_1 \in \mathbb{R}^{\Omega}$, we have that $\underline{\pi}\left(X_1\right) + \underline{\pi}\left(Y_1\right) \leq \underline{\pi}\left(X_1 + Y_1\right)$, with $i, j \in \{1, \ldots, K\}$, where equality holds only if X_1 and Y_1 are *comonotone*[4].

Theorem 4.1 (Chateauneuf et al., 1996). Under assumptions (A1)–(A3), for all $X_1 \in \mathbb{R}^{\Omega}$ there exists a unique convex capacity ν such that the *global* lower pricing rule $\underline{\pi}$ can be expressed as a discounted Choquet expectation of the payoff with respect to ν:

$$\underline{\pi}\left(X_1\right) = (1 + r)^{-1} \oint_{\Omega} X_1 \, \mathrm{d}\nu = (1 + r)^{-1} \mathbb{C}_{\nu}\left(X_1\right). \tag{29}$$

We stress that the Choquet integral with respect to a 2-monotone capacity is equivalent to the lower expectation with respect to the set of probability measures in $\mathbf{core}(\nu)$, hence the bid price is equivalently computed as:

$$\underline{X}_0 = \underline{\pi}\left(X_1\right) = (1 + r)^{-1} \min_{P \in \mathbf{core}(\nu)} \mathbb{E}_P\left(X_1\right). \tag{30}$$

As we already pointed out, the same model can be set up from the *upper* point of view with respect to a concave capacity, replacing assumption (A3) with *sub-additivity* property (that is the version in Chateauneuf et al., 1996). Since it is the dual function of a convex capacity, for each $X_1 \in \mathbb{R}^{\Omega}$, we can compute the ask price $\overline{X}_0$ as a discounted Choquet integral with respect to the conjugate concave capacity, which can be expressed in terms of an upper pricing rule:

$$\overline{X}_0 = \overline{\pi}\left(X_1\right) = -\underline{\pi}(-X_1). \tag{31}$$

The approach of Chateauneuf et al. (1996) characterizes a lower pricing rule already defined on the whole $\mathbb{R}^{\Omega}$. If we refer to the K fixed risky assets and identify the payoff of the risk-free bond B_1 with a 0-th asset, and $-B_1$ with a $(K+1)$-th asset, we have a finite set of payoffs $\mathcal{G} = \{S_1^{(0)}, \ldots, S_1^{(K+1)}\}$. In this case, we have a lower price assessment $\underline{\pi} : \mathcal{G} \to \mathbb{R}$ such that $\underline{\pi}\left(S_1^{(k)}\right) = \underline{S}_0^{(k)}$, for $k = 1, \ldots, K$, $\underline{\pi}\left(S_1^{(0)}\right) = 1$ and $\underline{\pi}\left(S_1^{(K+1)}\right) = -1$. Now, our goal is to find a lower pricing rule $\underline{\pi}' : \mathbb{R}^{\Omega} \to \mathbb{R}$ that extends $\underline{\pi}$ and can be expressed as a discounted Choquet expectation. This problem can be tackled in the framework of belief functions by relying on results given in Coletti et al. (2020).

If there exists a belief function $Bel : \mathcal{F} \to [0, 1]$ such that, for $k = 0, \ldots, K + 1$, the lower price assessment is defined as the discounted Choquet expectation with respect to Bel, that is it satisfies:

$$\underline{\pi}\left(S_1^{(k)}\right) = \mathbb{C}_{Bel}\left(\tilde{S}_1^{(k)}\right), \tag{32}$$

---

[4] Two assets $X_1, Y_1 \in \mathbb{R}^{\Omega}$ are comonotone if they vary in the same way: $\forall \omega, \omega' \in \Omega$, $[X_1(\omega) - X_1(\omega')][Y_1(\omega) - Y_1(\omega')] \geq 0$.

then the lower price assessment is called *CBel-coherent*. As usual, $\tilde{S}_1^{(k)}$ denotes the discounted payoff, for $k = 0, \ldots, K+1$.

Theorem 4.2 (Coletti et al., 2020). For a finite $\mathcal{G}$ defined as above the following statements are equivalent:

(i) $\underline{\pi}$ is a CBel-coherent price assessment;

(ii) $\underline{\pi}$ avoids *CBel-Dutch book* opportunities: for every $\boldsymbol{\lambda} \in \mathbb{R}^{K+2}$, the following condition holds:

$$\max_{B \in \mathcal{F} \setminus \{\emptyset\}} \sum_{k=0}^{K+1} \lambda_k \left(\min_{i \in B} \tilde{S}_1^{(k)}(i) - \underline{\pi}\left(S_1^{(k)}\right) \right) \geq 0. \tag{33}$$

Condition (33) assures that there cannot be a portfolio that leads to a sure loss, defined under *partially resolving uncertainty* (Jaffray, 1989), i.e., working over $\mathcal{F} \setminus \{\emptyset\}$.

The no-Dutch book condition in the setting of belief functions in Equation (33) is weaker than the classical no-Dutch book condition in Equation (3) since in the latter case we are working under *completely resolving uncertainty*. Completely resolving uncertainty is the common assumption of the classical Dutch-book condition (Equation (3)) and requires that, once uncertainty is resolved, the knowledge of the true state $i \in \Omega$ will be acquired. Conversely, under *partially resolving uncertainty* we assume that, when uncertainty is resolved, we may acquire the information that an event B has occurred but we may not identify the state $i \in B$ that turns out to be true. In particular, condition (33) considers a systematically pessimistic behavior as, for every $k = 0, \ldots, K+1$, we take the minimum payoff given by all $i \in B$, defined as $\min_{i \in B} \tilde{S}_1^{(k)}(i)$. We notice that Theorem 4.2 does not guarantee the uniqueness of Bel and, so, of the lower pricing ruler $\underline{\pi'}$ extending $\underline{\pi}$. Nevertheless, every such extension satisfies conditions (A1)–(A3) introduced before.

We finally get back to the trinomial market model. In Cinfrignini et al. (2021) it is proved that the lower probability $\underline{Q}$ of the set $\mathcal{Q}$ of equivalent martingale measures, computed as $\underline{Q}(A) = \min_{Q \in \mathrm{cl}(\mathcal{Q})} Q(A)$, $\forall A \in \mathcal{F}$, is a belief function. Therefore, this would suggest to define a lower pricing rule as the discounted Choquet expectation with respect to $\underline{Q}$. Unfortunately, a situation analogous to Example 3.1 occurs since the closure $\mathrm{cl}(\mathcal{Q})$ does not coincide with $\mathbf{core}(\underline{Q})$. Thus the discounted Choquet expectation does not coincide with the lower expectation computed with respect to $\mathcal{Q}$, and so the two approaches lead to different results.

We also notice that, still referring to the trinomial model, by considering the set $\mathcal{G} = \{B_1, S_1, -B_1\}$, with the lower pricing assessment defined as $\underline{\pi}(B_1) = 1$, $\underline{\pi}(S_1) = S_0$, and $\underline{\pi}(-B_1) = -1$, we get that the no-Dutch book condition in *(ii)* of Theorem 4.2 holds. It is actually possible to show that $\underline{\pi}$ can be extended by a discounted Choquet expectation functional computed with respect to a non-additive belief function Bel which, however, must be different from $\underline{Q}$.

5 Conclusion

In this paper we have presented a survey on classical pricing theory and we focused on markets with frictions in the form of bid-ask spreads. Frictions are largely proved to exists and are studied in order to embody them into price models. Then, after having introduced non-additive measures and the Choquet expectation, we recalled the properties characterizing a global lower pricing rule defined as the discounted Choquet expectation with respect to a convex capacity (Chateauneuf et al., 1996). Then, referring to a finite set of payoffs, we showed a condition that guarantees the representation of lower prices as discounted Choquet expectation with respect to a belief function. The latter condition is in the form of no-Dutch book under partially resolving uncertainty. We also showed that the lower envelope of equivalent martingale measures in the trinomial model does not produce sharp lower prices, with respect to the class of martingale measures, if used to compute discounted Choquet expectations (see Cinfrignini et al., 2021). Nevertheless, the lower prices of fixed securities satisfy the generalized no-Dutch book condition given in Coletti et al. (2020).

References

Acciaio B., Beiglbck M., Penkner F. and Schachermayer W. (2016), A model-free version of the fundamental theorem of asset pricing and the super-replication theorem, *Mathematical Finance*, 26 (2), 233-251.

Allingham M. (1991), *Arbitrage. Elements of Financial Economics*, Palgrave Macmillan, London.

Amihud Y. and Mendelson H. (1986), Asset pricing and the bid-ask spread, *Journal of Financial Economics*, 17 (2), 223-249.

Amihud Y. and Mendelson H. (1991), Liquidity, Maturity, and the Yields on U.S. Treasury Securities, *The Journal of Finance*, 46 (4), 1411-1425.

Bensaid B., Lesne J-P., Pags H. and Scheinkman J. (1992), Derivative asset pricing with transaction costs, *Mathematical Finance*, 2 (2), 63-86.

Bertsimas D., Kogan L. and Lo A. W. (2001), Hedging Derivative Securities and Incomplete Markets: An ϵ-Arbitrage Approach, *Operations Research*, 49 (3), 372397.

Capotorti A., Coletti G. and Vantaggi, B. (2008), Preferences Representable by a Lower Expectation: Some Characterizations, *Theory and Decision*, 64, 119146.

Cern A. (2009), *Mathematical Techniques in Finance: Tools for Incomplete Markets*. Princeton: Princeton University Press.

Cerreia-Vioglio S., Maccheroni F. and Marinacci M. (2015), PutCall Parity and market frictions, *Journal of Economic Theory*, 157 (C), 730-762.

Chateauneuf A. and Cornet B. (2022), Submodular financial markets with frictions, *Economic Theory*, 73, 721744.

Chateauneuf A. and Jaffray J.-Y. (1989), Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion, *Mathematical Social Sciences*, 17 (3), 263-283.

Chateauneuf A., Kast R. and Lapied A. (1996), Choquet Pricing For Financial Markets With Frictions, *Mathematical Finance*, 6 (3), 323-330.

Chen Z. and Kulperger R. (2006), Minimax pricing and Choquet pricing, *Insurance: Mathematics and Economics*, 38 (3), 518-528.

Cinfrignini A., Petturiti D. and Vantaggi B. (2021), Envelopes of equivalent martingale measures and a generalized no-arbitrage principle in a finite setting, *arXiv:2107.01240*

Coletti G., Petturiti D. and Vantaggi B. (2016), Conditional belief functions as lower envelopes of conditional probabilities in a finite setting, *Information Sciences*, 339, 64-84.

Coletti G., Petturiti D. and Vantaggi B. (2020), A Dutch book coherence condition for conditional completely alternating Choquet expectations, *Bollettino dell'Unione Matematica Italiana*, 13, 585593.

Cox J.C., Ross S.A. and Rubinstein M. (1979), Option pricing: A simplified approach, *Journal of Financial Economics*, 7, 229-263.

Cozman F.G. (2000), Credal networks, *Artificial Intelligence*, 120 (2), 199-233.

Dempster A.P. (1967), Upper and Lower Probabilities Induced by a Multivalued Mapping, *The Annals of Mathematical Statistics*, 38 (2), 325-339.

Dybvig P.H. and Ross S. (1989), *Arbitrage*, in Eatwell J., Milgate M. and Newman P., *Finance*. Palgrave Macmillan, London.

Erreygers A., De Bock J., de Cooman G. and Van Camp A. (2019), Optimal control of a linear system subject to partially specified input noise, *International Journal of Robust and Nonlinear Control*, 29, 38923914.

Etner J., Jeleva M. and Tallon J-M. (2012), Decision Theory Under Ambiguity, *Journal of Economic Surveys*, 26 (2), 234-270.

Ghirardato P. and Marinacci M. (2001), Risk, Ambiguity, and the Separation of Utility and Beliefs, *Mathematics of Operations Research*, 26 (4), 864890.

Gilboa I. and Marinacci M. (2011), Ambiguity and the Bayesian Paradigm, *Working Papers 379*, IGIER, Bocconi University.

Gilboa I. and Schmeidler D. (1989), Maxmin expected utility with non-unique prior, *Journal of Mathematical Economics*, 18 (2), 141-153.

Gilboa I. and Schmeidler D. (1994), Additive Representations of Non-Additive Measures and the Choquet Integral, *Annals of Operations Research*, 52 (1), 43-65.

Jaffray J-Y. (1989), Coherent Bets Under Partially Resolving Uncertainty and Belief Functions, *Theory and Decision*, 26 (2), 99-105.

Jouini E. and Kallal H. (1995), Martingales and Arbitrage in Securities Markets with Transaction Costs, *Journal of Economic Theory*, 66, 178-197.

Pascucci A. and Runggaldier W.J. (2011), *Financial Mathematics. Theory and problems for multi-period models*, Springer Verlag, Italia, Series UNITEXT.

Petturiti D. and Vantaggi B. (2020), Modeling agent's conditional preferences under objective ambiguity in Dempster-Shafer theory, *International Journal of Approximate Reasoning*, 119, 151-176.

Petturiti D. and Vantaggi B. (2022), Conditional decisions under objective and subjective ambiguity in Dempster-Shafer theory, *Fuzzy Sets and Systems*, 447, 155-181.

Pliska S. (1997), *Introduction to Mathematical Finance: Discrete Time Models*, (2nd ed.), Malden: Blackwell, Oxford.

Runggaldier W. (2006), Portfolio optimization in discrete time, *Accademia delle Scienze dell'Istituto di Bologna*.

Schervish M.J., Seidenfeld T. and Kadane J.B. (2008), The Fundamental Theorems of Prevision and Asset Pricing, *International Journal of Approximate Reasoning*, 49 (1), 148-158.

Schmeidler D. (1986), Integral representation without additivity, *Proceedings of the American Mathematical Society*, 97 (2), 255261.

Schmeidler D. (1989), Subjective Probability and Expected Utility without Additivity, *Econometrica*, 57 (3), 571587.

Shafer G. (1976), *A Mathematical Theory of Evidence*. Princeton: Princeton University Press.

T'Joens N., De Bock J. and de Cooman G. (2021), Game-theoretic upper expectations for discrete-time finite-state uncertain processes, *Journal of Mathematical Analysis and Applications*, 504 (2).

Walley P. (1991), *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall, London.

SAPIENZA
UNIVERSITÀ EDITRICE

ANNALI DEL DIPARTIMENTO DI METODI E MODELLI
PER L'ECONOMIA IL TERRITORIO E LA FINANZA
2022 | pp. 29-44

E-ISSN 2611-6634 | ISSN 2385-0825
DOI: 10.13133/2611-6634/1393

Research article

Lorenzo Giammei[*]

AN INTEGRATED APPROACH TO CAUSALITY: THE ROLE OF CAUSAL GRAPHS

Abstract

Causal questions are central for most biomedical and social science studies. The main frameworks that allow the analysis of causal relations are Potential Outcomes and Causal Graphs. The approaches have often been compared, contrasting their relative strengths. This paper evaluates the implications of merging the two methodologies in an integrated approach. In particular, we assess how the limits of one can be compensated by the solutions provided by the other. The outlined approach employs causal graphs to discover and formalize a causal model that is then used as a guide to implementing potential outcomes identification strategies. The integrated approach could be beneficial to both frameworks. The assumptions of potential outcome methods can be assessed directly from a causal graph even in high dimensional contexts, thus making the obtained causal estimates more reliable. On the other hand, causal graphs can benefit from the several ad hoc identification strategies that have been developed in the potential outcomes literature.

Keywords: causality, causal graphs, potential outcomes.

---

[*] Department of Methods and Models for Economics, Territory and Finance, Sapienza University of Rome

1 Introduction

The study of cause and effect relations motivates most research in social, demographic and health sciences. Investigating causality usually means assessing if and how a certain intervention, often called treatment, affects an outcome of interest. The early work of Neyman and Iwaszkiewicz (1935), Fisher (1949) and Cox (1958) in the field of randomized experiments constituted a first step towards a rigorous analysis of causality. Based on these studies Rubin (1974) formalizes one of the most relevant approaches to causality: the *Potential Outcomes* (PO) framework. The framework has then been enriched with many contributions that proposed new methods and applications (Imbens and Rubin, 2015; Rosenbaum, 2018). PO have a strong connection with economics since its early stages as its concepts are rooted in the work of Tinbergen (1930) and Haavelmo (1943). PO methods are now widely applied in statistics and economics and many econometric textbooks solely rely on this approach (Angrist and Pischke, 2008; Imbens and Rubin, 2015).

The other main approach to deal with causality is the *Causal Graph* framework. Note that causal graphs, also called Causal Bayesian Networks or Causal Diagrams, can be seen as part of a wider model called structural causal model (SCM) (Pearl, 2000). In a SCM the causal graph is also associated to a set of equations that describe causal relations between the nodes of the graph. Here we will however only focus on the causal graph component, that is sufficient for answering causal queries concerning the effect of interventions. Causal graphs are described in Pearl (2000) and share some elements with the previous work on path diagrams in Wright (1921). The framework has been subsequently developed and enriched with several contributions that extended its applicability and strengthened its results (Pearl, 2000; Tian and Pearl, 2002; Bareinboim and Pearl, 2016; Huang and Valtorta, 2006). Causal graphs are now frequently used in epidemiology, computer science and some social sciences, though they are still uncommon in economics.

The relative advantages of the two frameworks have been recently reviewed and compared in Imbens (2020) and Hnermund and Bareinboim (2019). Both papers show some specific causal problems where one approach is more appropriate than the other and vice-versa, thus revealing that, at least in part, the two are complementary and could benefit from each other. The idea of an integrated approach also starts to appear in some causal inference textbooks, such as Morgan and Winship (2015) and Cunningham (2021), however integrated applications are still very rare in practice.

In this paper, we assess how PO and Causal Graphs can be combined and the implications of carrying out such an approach. The basic ideas of the frameworks will be described focusing on when the limits of one way of proceeding are compensated by the other. Particular attention will be put on causal discovery techniques, a resource that is often overlooked when comparing PO and causal graphs. Throughout the paper, we provide some basic examples in which the combination of the frameworks can improve the results' quality and reliability.

Section 2 will outline the PO framework, its main assumptions, results and limits. Section 3 is instead devoted to Causal Graphs. The basic terminology is presented and the principal features are described with the help of some examples. Then we show how causal effect estimation can be performed from causal graphs and how the process can be integrated with PO methods. Finally, in section 4 we introduce the concept of causal discovery; we explain how structural learning algorithms work and why they can be valuable for causal effect estimation.

2 Potential Outcomes

The Potential Outcomes framework originates from the work of Splawa-Neyman et al. (1990) and Rubin (1974) on randomized controlled trials (RCT). The name of the framework comes from its peculiar notation $Y_i(t)$ that denotes the *potential outcome* for unit i when receiving the treatment level $T = t$. In the case of a binary treatment, T takes value 1 if unit i is treated and 0 otherwise. Accordingly, $Y_i(1)$ represents the potential outcome we would observe for unit i if it was treated and $Y_i(0)$ the potential outcome if unit i was a control. The causal effect of T on Y can therefore be computed by comparing summary statistics of the potential outcomes distribution. The resulting causal estimate is usually called the average treatment effect (ATE) and can be expressed in different ways, such as:

$$ATE = E[Y_i(1) - Y_i(0)] \qquad \text{or} \qquad ATE = \frac{E[Y_i(1)]}{E[Y_i(0)]}.$$

However, the ATE cannot be estimated directly from data since only one of the potential outcomes is observed for each unit i. Units receive only one level of treatment, creating a missing data problem. This is sometimes referred to as the fundamental problem of causal inference (Holland, 1986). PO literature contributed to answering this problem in the context of randomized experiments. In this setting, treatment is assigned randomly to the units of the sample, thus rendering T independent of the potential outcomes (by symbol $T_i \perp\!\!\!\perp (Y_i(0), Y_i(1))$).

This scenario, together with the assumption that there is no interference between units (SUTVA)(Imbens and Rubin, 2015), ensure that an unbiased estimate of the ATE can be obtained by computing the difference:

$$\bar{Y_t} - \bar{Y_c}, \qquad \text{with} \qquad \bar{Y_t} = \frac{1}{N_t} \sum_{i:T_i=1} Y_i \qquad \text{and} \qquad \bar{Y_c} = \frac{1}{N_c} \sum_{i:T_i=0} Y_i.$$

The indexes $i : T_i = t$ indicate to sum over the units that received a certain treatment level; N_t and N_c denote respectively the number of treated and control units.

The PO framework also provides several solutions to deal with non-experimental or observational data. What usually prevents observational data from being treated as experimental data is the presence of *confounders*. Confounders are variables that affect both the treatment and the outcome and can lead to biased causal estimates if not adequately accounted for. The concern worsens when confounders are unobserved since, in this situation, treatment effects could be impossible to identify.

PO methods that deal with observational data aim at emulating an experimental context under specific assumptions. One of these, that tackles directly the problem of confounders, is called *unconfoundedness* or *ignorability* and can be defined by symbols as $T_i \perp\!\!\!\perp (Y_i(0), Y_i(1))|X_i$, where X_i is a set of pre-treatment covariates. Unconfoundedness states that the treatment T_i is independent of the potential outcomes, given a set of pre-treatment variables X_i. The condition allows estimating the ATE as:

$$ATE = E[Y_i(1) - Y_i(0)] = E[E[Y_i|T_i = 1, X_i] - E[Y_i|T_i = 0, X_i]]. \tag{1}$$

The formula in Equation 1 is also called *adjusting for X* and as long as unconfoundedness holds, it ensures an unbiased estimation of the ATE in the presence of confounders. Adjustment can be performed through various methods, including regression, matching and inverse probability weighting.

Another PO method to derive causal estimates from observational data is the instrumental variable (IV) strategy (Angrist, 1990). In this context, there is an unobserved variable U, which violates the unconfoundedness assumption for the effect of T on Y. Since U is unobserved, it is impossible to adjust for it in order to obtain unbiased estimates. However, if the treatment T is affected by another variable Z, it is still possible to estimate a causal effect, under an assumption called *exclusion restriction*. The assumption can be expressed as:

$$Y_i(z,t) = Y_i(z',t) \qquad \text{for all } z, z',$$

imposing that potential outcomes do not vary with Z. PO literature refers to variables that satisfy the exclusion restriction as instrumental variables. However, exclusion restriction and unconfoundedness cannot be tested, and they are usually motivated by background theory concerning the causal relations between variables. This implies that justifying them becomes difficult if a priori knowledge is missing. Moreover, as the number of variables in the model increases, assessing the two assumptions' validity turns out to be a challenging task.

The PO framework includes many more identification strategies, such as difference-in-differences, regression discontinuity and synthetic control. For a review of the newest techniques, see Athey and Imbens (2017). These methods provide solutions to very specific causal problems and usually impose additional functional-forms restrictions on probability distributions, such as linearity, monotonicity or additivity.

3 Causal Graphs

In this section, the Causal Graph framework will be described. First, we will introduce the basic terminology of graphs and the main elements of causal graph theory. Next, we will show how interventions are represented in the framework and how causal effects can be estimated employing graphs.

3.1 Terminology and basic concepts

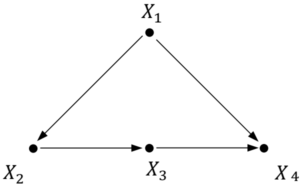
A *graph* $G = (V, E)$ is a collection of vertices or nodes V and edges E. The edges can be directed or indirected. An edge that goes from a vertex V_i to another vertex V_j is a *directed edge*. Conversely, an edge without such orientation is an *undirected edge*. A graph that only contains directed edges is called a *directed graph*. When two nodes are connected by an edge, they are called *adjacent* nodes. If each pair of nodes belonging to V is connected by an edge, the graph is called a *complete graph*. Conversely, if none of the pairs is adjacent, the graph is an *empty graph*.

A sequence of connected edges that starts from a node V_i and ends with node V_j, regardless of the directions of the edges, is called a *path*. In a *directed path* all the edges are oriented in the same direction along the path. A directed path, starting from V_j and ending in V_i, with $V_j = V_i$ is a *cycle*.

A directed graph that contains no cycles is also called a *directed acyclic graph* (DAG) (Pearl, 2000). In the context of causal graphs, DAGs are employed to represent causal structures. The vertices of the DAG represent random variables, and its edges describe the causal relations between them. We will refer to variables and vertices in a DAG interchangeably from now on.

Consider the graph G in Figure 1. All the edges in the graph are directed, and they form no cycles; the graph is, therefore, a DAG. G describes the multivariate causal relations between a set of four random variables **X**. The terminology of kinship is often used to indicate relationships between nodes according to the graph's structure. Since the DAG contains a directed edge going from X_1 to X_2, X_1 is called a *parent* of X_2 and the latter is a *child* of X_1. The path p along the ordered sequence of nodes (X_1, X_2, X_3, X_4) is a directed path since all the edges are oriented in the same direction along the path. X_1 is called an *ancestor* of each node belonging to $\{X_2, X_3, X_4\}$ since it precedes them in p and the vertices in $\{X_2, X_3, X_4\}$ are *descendants* of X_1. Given that the edges are carriers of causal information, we can also say that X_1 is a direct cause of X_2 and X_4. The same is true for every ordered pair of random variables (X_i, X_j) connected by a directed edge that goes from X_i to X_j in the DAG.

Figure 1. A simple DAG



Every causal graph also consists of a joint probability distribution $P(\mathbf{X})$ over the variables described by the DAG. This distribution can be factorized according to the structure of the DAG as:

$$P(x_1, , x_n) = \prod_i P(x_i|pa_i), \tag{2}$$

where pa_i indicate the parent set of variable X_i. The factorization implies that given a DAG G with node set **X**, for each variable $X_i \in \mathbf{X}$, its parent set PA_i selected according to the structure of G, is sufficient for determining the probability of X_i. If a probability function P admits the factorization of Equation 2 relative to a DAG G, then G is said to satisfy the *causal Markov condition* and P is said to be *Markov* relative to G.

The edges of a DAG can assume specific configurations that provide additional information regarding the independence relations among variables of the model. Given the ordered triplet of nodes (X_i, X_j, X_k), if two directed edges goes from X_i and X_k to X_j but X_i and X_k are not adjacent, then

X_j is called a *collider* (or unshielded collider) in the ordered triplet. Colliders are also referred to as *non-emitting nodes*. Conversely, given a path p, vertices belonging to p with at least an outgoing edge directed towards other adjacent nodes in p are called *emitting nodes*. An example of a configuration that only contains emitting nodes is when a directed edge goes from node X_i to node X_j, and another directed edge goes from X_j to a third node X_k. This configuration is called *chain*.

A DAG encodes information concerning conditional independence among the variables it represents through a criterion called *d-separation*. Consider a DAG G with node set **X**, a pair of nodes $\{X_i, X_j\}$ belonging to **X** with $X_i \neq X_j$ and a set of nodes $S \subset \mathbf{X}$ not containing X_i and X_j. A path p between X_i and X_j is said to be blocked by a set **S** in G, if either:

1. p contains at least one arrow-emitting node that belongs to **S**;
2. p contains at least a collision node that does not belong to **S** and has no descendent in **S**.

Two nodes X_i and X_j are said to be *d-separated* given a set **S** if all the paths between the nodes are blocked by **S**. When two nodes X_i and X_j are d-separated by a set **S**, then X_i is independent of X_j conditional on **S**. Note that two nodes can also be d-separated conditioning on an empty set if all the paths between them contain at least a collider or its descendants. In this case, the variables represented by the nodes are said to be *marginally independent*.

3.2 Causal graph analysis at interventional level

Causal graphs allow estimating the effect of interventions, or in other words, the effect of forcing a variable to take a certain value by an external action. Pearl (2000) introduces the *do-operator* $do(X = x)$, a notation to indicate that a variable X is forced by intervention to take value x. In order to be coherent with the terminology defined in Section 2 for the PO framework, we will refer to the effect of a treatment variable T on an outcome variable Y. The do-operator allows writing $P(Y|do(T = t))$ to denote the distribution of Y given an intervention that sets $T = t$. This is different from $P(Y|T = t)$ that instead represents the observational distribution of Y given $T = t$. The causal effect of T on Y can thus be obtained by comparing the quantity $P(Y|do(T = t))$ for different values of t, similarly to what is done in the PO framework where instead $Y(t)$ was the quantity of interest. However, when dealing with non-experimental data, causal effects cannot be estimated directly from data since the interventional distribution of Y is not an observed quantity.

One of the critical contributions of causal graphs is that their structure can serve as a guide to express interventional distributions in terms of observational quantities, thus making it possible to estimate causal effects. This is a crucial result since conditional distributions, such as $P(Y|T = t)$, can be directly computed in a non-experimental context through the joint probability distribution associated with the DAG.

A graphical condition called *back-door criterion* can be applied to a given causal graph to test if a subset of its nodes **S** is sufficient for identifying $P(Y|do(T = t))$ from observational data. A set of variables $\mathbf{S} \subseteq \mathbf{X}$ satisfies the back-door criterion relative to a graph G with node set **X**, a treatment variable $T \in \mathbf{X}$ and an outcome variable $Y \in \mathbf{X}$ if:

1. no node in **S** is a descendant of T;

   and

2. **S** blocks all the paths between T and Y that contain a directed edge pointing towards T.

If the back-door criterion is satisfied by a set **S**, then interventional quantities can be expressed through observational ones as follows:

$$P(y|do(T=t)) = \sum_{\mathbf{S}} P(y|t, \mathbf{s})P(\mathbf{s}). \tag{3}$$

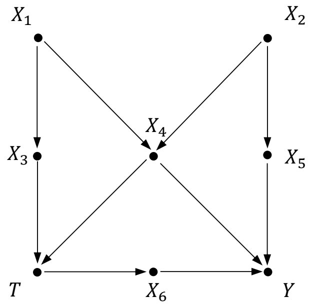
The formula used to compute the interventional probability distribution of the outcome in Equation 3 is also known as adjusting for **S**. Summary statistics of the interventional distributions can then be compared to compute the ATE. Obtaining an adjustment set **S** through the back-door criterion also ensures that **S** satisfies the unconfoundedness condition for estimating the effect of T on Y. Therefore, performing a matching procedure (Imbens and Rubin, 2015) by balancing the variable set **S**, would ensure obtaining unbiased estimates of the ATE. This is an example of how Causal graphs can be used as guides for assessing and justifying the assumptions some PO methods require.

Suppose we are interested in estimating $P(y|do(T=t))$ given a causal model represented by the DAG in Figure 2 (Pearl, 2000), with node set $\{\mathbf{X}, T, Y\}$ and a joint probability distribution $P(\mathbf{X}, T, Y)$. The knowledge of the DAG allows the application of the back-door criterion to select an adjustment set for causal effect estimation. The procedure reveals that adjusting for the set $\{X_3, X_4\}$ or $\{X_4, X_5\}$ ensures unbiased estimates of $P(y|do(T=t))$. Conversely, performing the adjustment procedure on a set $\mathbf{S} = \{X_4\}$ would produce biased estimates, since the set does not block all the back-door paths between X and Y.

Let us now consider the graph in Figure 3. The DAG shows the presence of two unobserved or latent variables U_1 and U_2. The two nodes are denoted by a circle rather than a solid dot to indicate the variables are not observed. Even in the presence of latent variables, we can resort to the back-door criterion to assess if an adjustment set to estimate the effect of T on Y exists. In this scenario, we are particularly interested in checking if some of the sets that satisfy the back-door criterion are composed only by observed variables. In this situation, adjusting for $\{X_1\}$ would open the back-door path along the ordered tuple (T, U_1, X_1, U_2, Y), thus producing a biased estimate of the effect of T on Y. Conditioning on the empty set provides instead unbiased estimates of the causal effect, since the colliding path over the ordered triplet (U_1, X_1, U_2) is blocked as long as we do not condition on X_1.

The bias introduced by conditioning on X_1 is also called $M - bias$, and it constitutes a solid motivating argument for employing causal graphs. Generally, the PO literature suggests to condition on all the observed pre-treatment variables in order to improve the quality of causal estimates (Imbens and Rubin, 2015). However, in this scenario and similar ones, conditioning on the observed variables leads instead to worse causal estimates, and causal graphs provide a rule, namely the back-door criterion, to avoid this sort of bias. For a review on how conditioning can affect causal estimates, given different contexts represented by causal graphs, see Cinelli et al. (2020).

Figure 2. A DAG describing causal relations among a set of variables **X***, a treatment T and an outcome Y*



The back-door criterion is not the only strategy that can be employed to estimate causal effects from a causal graph. Pearl (2000) describes a specific graphical configuration that allows causal effect identification, even when back-door adjustment is not feasible. The condition is called *front-door criterion* and states that given a DAG G with node set **X**, a set **S** $\subset$ **X** satisfies the front-door criterion for the effect of T on Y, both belonging to **X**, if:

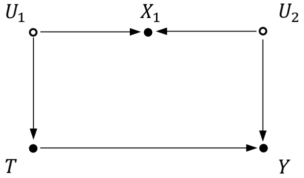
1. **S** intercepts all directed paths from T to Y;
2. all the back-door paths from T to **S** are blocked;
3. all the back-door paths from **S** to Y are blocked by T.

If a set **S** that satisfies the front-door criterion for the effect of T on Y exists and $P(t, s) > 0$, then the causal effect of T on Y can be computed with the formula

$$P(y|do(T = t)) = \sum_s P(s|t) \sum_{t'} P(y|t', s)P(t'). \tag{4}$$

Combined and iterative use of back-door and front-door criterion constitute the building block to identify causal effects on complex DAGs. Pearl (2000) describes a set of rules based on the two criteria, also called *do-calculus*, that allows expressing interventional distributions in terms of observational distributions only, in an automated way. The procedure has been proved to be sound and complete, meaning that an algorithmic iteration of the rules of do-calculus always returns a solution for the identification of causal effects, if such solution exists (Pearl, 2000; Tian and Pearl, 2002; Huang and Valtorta, 2006).

Figure 3. A DAG with unobserved confounders



4 Causal discovery

Causal graphs are powerful models to describe the causal structure of a set of random variables. Moreover, they constitute a guide for selecting an identification strategy to estimate causal effects. However, the setting considered here always assumed a complete knowledge of the causal diagram.

Suppose we want to investigate the causal effect of a treatment variable T on an outcome variable Y from a dataset $D(\mathbf{X}, T, Y)$ where $\mathbf{X}$ is a set of other covariates. We also assume the existence of an unknown underlying causal model described by a DAG $G(V, E)$ and a joint probability distribution $P(V)$, from which $D(\mathbf{X}, T, Y)$ has been sampled. In order to obtain an unbiased estimate of $P(Y|do(T = t))$ we therefore study if it is possible to *learn* a causal graph from $D(\mathbf{X}, T, Y)$. In order to estimate the structure of the causal DAG, *structural learning algorithms* have been developed. These algorithms take a dataset as an input and, under a set of assumptions, recover a DAG and the associated joint probability distribution. This process is known as *causal discovery* (Spirtes et al., 2000). Structural learning algorithms can be divided in three families: *constraint-based* algorithms, *score-based* algorithms and *hybrid* algorithms.

Constraint-based algorithms learn the graph's structure via conditional independence statements emerging from data. They usually start with a complete graph, and then if two variables turn out to be marginally or conditionally independent, the edge connecting them is deleted. This procedure is repeated iteratively until a stopping criterion is satisfied.

Score-based algorithms rely on a given score function that measures how well a certain DAG describes a dataset. These algorithms usually begin by computing the score of an initial graph. The diagram is then modified by introducing, deleting or reversing edges, and its score is computed again for each modification. The graph recording the best score at the end of the procedure is retained as the algorithm's output.

Hybrid algorithms aim to exploit the advantages of score-based and constraint-based algorithms by merging them in a single procedure. Generally, they begin with a *restrict* phase where the parents of each node are selected through tests of conditional independence, similarly to what happens in constraint-based algorithms. The second phase is called *maximize* and consists in selecting a DAG in the restricted DAG family outlined by phase one by optimizing a given score function. Hybrid algorithms include the Max-Min Hill Climbing (Tsamardinos et al., 2006) and $H2PC$ (Gasse et al., 2014).

Once the graph is learnt, a joint probability distribution over the nodes of the graph can be obtained through maximum likelihood estimation. This phase usually involves computing maximum likelihood estimates subject to the independence constraints encoded in the graph. Estimates can be retrieved in the case of discrete variables or when dealing with continuous variables under the assumption of linearity (Spirtes et al., 2000).

The section will continue with a description of the assumptions that structural algorithms usually require. We will then explain how different algorithms work and show the functioning of two representative procedures.

4.1 Common assumptions and background knowledge

The assumptions of causal discovery algorithms usually focus on the relation between the causal graph and the distribution employed to learn it. A usually required assumption is *faithfulness*. A graph G faithfully represents a dataset D, if and only if all and only the list of d-separations emerging from D are true in G. This ensures an exact correspondence between conditional independence relations of the distribution from which the graph is learnt and those entailed by the causal Markov condition applied to G. Another key assumption for learning algorithms is *causal sufficiency*. The assumption states that a given set of variables **X** is causally sufficient for a population if and only if in the population every common cause of any two or more variables belonging to **X** is in **X** or has the same value for all units in the population. Implementing a constraint-based algorithm also requires making statistical decisions concerning how to assess conditional independence. Several tests can be employed to check if conditional independence holds, and violations of the assumptions required by the tests can generate unreliable independence statements. For a review of the implications of choosing a given independence test and what happens when the required assumptions do not hold, see Spirtes et al. (2000).

Structural learning algorithms are usually employed when information concerning the causal graph is not available. However, in practice it is common to deal with scenarios where the knowledge of the causal graph is partial. This incomplete knowledge can be introduced in structural learning procedures by imposing constraints on the structure of the obtained network. For example, if is known that a variable X_i cannot cause a second variable X_j, the directed edge that goes from X_i to X_j is forced to be absent. Note that this constraint does not imply the presence or absence of a directed edge going from X_j to X_i. Conversely, if background knowledge suggests that X_i affects X_j, a directed edge from X_i to X_j can be imposed.

A consequence of including previous knowledge in the learning phase is that the graph is not entirely obtained through the information contained in the data. The constraints on the structure of the graph restrict the search space of the algorithms and often reduce both uncertainty and computational time.

4.2 Constraint-based algorithms

Constraint-based algorithms learn causal graphs from conditional independence relations contained in the data. They can usually take both discrete and linear continuous data as input: in the first case, the algorithm performs conditional independence tests on cell counts; in the latter, covariance matrices are used to test vanishing partial correlations. The obtained conditional independence statements, if possible, are then translated into graphical form according to the rules of d-separation. Constraint-based algorithm include the PC algorithm(Glymour et al., 1991), the IG algorithm (Verma and Pearl, 1990) and the most recent Grow-Shrink algorithm (Margaritis, 2003). All the algorithms share the idea of learning a graph from the independence structure of the data but employ different heuristics. Constraint-based algorithms generally assume causal sufficiency, namely observing all the common causes of two or more variables in the model. This is a strong assumption, difficult to achieve in observational contexts. Some constraint-based algorithms have been proposed to deal with models where causal sufficiency does not hold. One of the most used is the fast causal inference (FCI) algorithm (Spirtes et al., 2000). The algorithm is a variation of the PC algorithm and retrieves asymptotically correct causal structures in the presence of latent common causes, provided the observed distribution and the graph satisfy the faithfulness condition.

One of the most used algorithms in the constraint-based family is the PC algorithm. The procedure begins with a complete undirected graph, in which edges are progressively deleted when they describe relations between variables that are found to be conditionally independent. Faithfulness and causal sufficiency are assumed. A pseudocode of a recent variation of the algorithm, called *PC-stable* (Colombo and Maathuis, 2014) is displayed in Algorithm 1. In the original PC algorithm the obtained graph could be affected by the ordering of the variables in the dataset used to learn the graph. In the new version, instead, the ordering does not affect the results, thus the name PC-stable. The procedure begins by learning a graph containing only undirected edges from conditional independence statements retrieved from the dataset. Then the orientation of the edges is determined according to a set of graphical rules. In the pseudocode we will denote directed and undirected edges between to nodes X_i and X_j, respectively with the notation $X_i \rightarrow X_j$ and $X_i - X_j$. Moreover we will use $adj(X_i)$ to denote the set composed by the nodes adjacent to X_i and $\mathbf{X}\backslash\{X_i\}$ to indicate the variable set **X** excluding variable X_i.

Given a dataset $D = (\mathbf{X})$ describing a set of random variables $\mathbf{X} = \{X_1, ..., X_N\}$, the PC-stable algorithm begins by forming a complete undirected graph G over **X**. Then, step 5 stores the adjacency sets $adj(G, X_i)$ for each node X_i according to the current structure of G. Given an index l which start from 0 and increase at each iteration, the procedure checks if a set **S** of size l, that d-separates two nodes X_i and X_j, exists. Note that **S** must be formed by nodes belonging to $adj(G, X_i)$ obtained in step 5 and that the size of $adj(G, X_i)\backslash X_j$ must be greater or equal than l. If the procedure finds a set **S** of size l that makes X_i and X_j conditionally independent, the edge between them is deleted from G and **S** is retained. The procedure is repeated for every node pair (X_i, X_j) and for every possible size l **S** associated to it, until an **S** that ensures d-separation is found or every **S** of size l has been explored. The algorithm then increases l by a unit and repeats the procedure from step 5, until every pair of adjacent nodes (X_i, X_j) in G satisfies $|a(X_i)\backslash\{X_j\}| \leq l$. In other words, at each iteration,

```
Algorithm 1: PC-stable
   Input: A sample $D = (\mathbf{X})$ from a set of random variables $\mathbf{X} = \{X_1, ..., X_N\}$ and a chosen statistical test of conditional
          independence
   Output: A family of Markov-equivalent DAGs
 1 Form a complete undirected graph $G$ with vertex set $\{X_1, ..., X_N\}$;
 2 Set $l = -1$;
 3 repeat
 4     $l = l + 1$;
 5     forall vertices $X_i$ in $G$ do
 6         Set $a(X_i) = adj(G, X_i)$
 7     end
 8     repeat
 9         select a (new) adjacent pair of nodes $(X_i, X_j), i \neq j$ in $G$ such that $|a(X_i) \backslash X_j| \geq l$;
10         repeat
11             Choose a (new) set $\mathbf{S} \subseteq a(X_i) \backslash \{X_j\}$ of size $l$;
12             if the statistical test reveals that $X_i$ is conditionally independent from $X_j$ given $\boldsymbol{S}$ then
13                 delete the edge connecting the pair $(X_i, X_j)$ from $G$;
14                 set $\mathbf{S}_{X_i X_j} = \mathbf{S}$, denoting the set that separates $X_i$ and $X_j$
15             end
16         until $X_i$ and $X_j$ are no longer adjacent in G or all possible subsets S of size l have been considered;
17     until all pairs of adjacent nodes $(X_i, X_j), i \neq j$ in G such that $|a(X_i) \backslash \{X_j\}| \geq l$ have been considered;
18 until all pairs of adjacent nodes $(X_i, X_j)$ in G satisfy $|a(X_i) \backslash \{X_j\}| \leq l$;
19 foreach triplet $\{X_i, X_k, X_j\}$ such that $X_i$ is adjacent to $X_k$, the latter is adjacent to $X_j$, but the pair $\{X_i, X_j\}$ is not adjacent
     to $X_j$, if $X_k \notin \boldsymbol{S}_{X_i, X_j}$ do
20     orient $X_i - X_k - X_j$ with the colliding configuration $X_i \rightarrow X_k \leftarrow X_j$.
21 end
22 Set more arc directions by repeated application the following rules:
23 if $X_i$ is adjacent to $X_j$ and there is a directed edge from $X_i$ to $X_j$ then
24     replace $X_i$ - $X_j$ with $X_i \rightarrow X_j$
25 end
26 if there are two paths $X_i - X_k \rightarrow X_j$ and $X_i - X_l \rightarrow X_j$ and $X_k$ is not adjacent to $X_l$ and there is a directed edge from $X_i$ to
     $X_j$ then
27     replace $X_j$ - $X_k$ with $X_j \rightarrow X_k$
28 end
29 if $X_i$ and $X_k$ are not adjacent but $X_i \rightarrow X_j$ and $X_j$ - $X_k$ then
30     replace $X_j - X_k$ with $X_j \rightarrow X_k$
31 end
```

the structure of G is updated by removing edges between conditional independent variables. Once the undirected graph is obtained, steps 19-31 orient the edges according to specific edge configurations. The rules dictated by the algorithm ensure that cycles are not generated and avoid the creation of a new colliding configuration that would modify the conditional independence relations.

The output of the PC-stable algorithm is a *completed partially DAG* (CPDAG), a DAG where some of the edges are undirected. This kind of graph is used to represent a family of independence-equivalent DAGs. Regardless of how the undirected edges of the graph are oriented, the colliding configurations remain the same, thus ensuring that all the DAGs associated to a CPDAG encode the same conditional independencies. The output of the PC-stable algorithm is therefore coherent with the objective of translating the conditional independencies contained in the data into graphical form. Moreover, it has been proven that, if the assumptions hold, the results provided by the algorithm are sound and complete (Colombo and Maathuis, 2014).

4.3 Score-based algorithms

Score-based algorithms aim at recovering the graph structure from data by optimizing a score function. Generally, this kind of algorithm explores several graph structures and assigns a score to each of them; at the end of the procedure, the graph with the maximal score is retained. Score-based algorithms usually assume faithfulness as well as causal sufficiency. Algorithms belonging to this family include the *greedy search*, the *simulated annealing* and *genetic algorithms* (Russel and Norvig, 2009).

The *greedy search* is one of the most used score-based algorithms, and its steps are shown in the pseudocode of Algorithm 2. The procedure iteratively modifies the edges of an initial DAG, computes the score of each graph and retains the best-scoring structure. When the score does not increase with an iteration, the obtained graph is provided as the algorithm's output.

Algorithm 2: Greedy Search

```
   Input: A sample $D = (\mathbf{X})$ from a set of random variables $\mathbf{X} = \{X_1, ..., X_N\}$ a score function $\mathcal{F}(G, D)$
   Output: A DAG
1  Form an empty graph $G$ with vertex set $\{X_1, ..., X_N\}$;
2  Calculate the score of $G$ given $D$, $S_G = \mathcal{F}(G, D)$ ;
3  Set $S_{max} = S_G$ ;
4  Set $G_{max} = G$ ;
5  repeat
6      foreach possible edge addition, removal or inversion in $G_{max}$ that produces a modified DAG $G^\star$ do
7          compute $S_{G^\star} = \mathcal{F}(G^\star, D)$;
8          if $S_{G^\star} > S_{max}$ and $S_{G^\star} > S_G$ then
9              set $G = G^\star$ and $S_G = S_{G^\star}$
10         end
11     end
12     if $S_G > S_{max}$ then
13         set $S_{max} = S_G$ and $G_{max} = G$
14     end
15 until $S_{max}$ of current iteration is smaller then $S_{max}$ of previous iteration;
```

Given a dataset $D = (\mathbf{X})$ and a score function $\mathcal{F}(G, D)$, the algorithm first two steps consist in computing the score of an initial, usually empty, graph G with vertex set **X**. Next, the score of the graph is set as the maximal score S_{max} and the initial graph G is set as the best-scoring DAG G_{max}. In step 6, the best-scoring DAG is modified by deleting, adding or inverting an edge, thus generating a new DAG $G^\star$. The score of $G^\star$ is computed and, if it is greater than the best score of the iteration S_G and greater than the absolute best score S_{max}, then $G^\star$ becomes the new best score of the iteration S_G. All the possible modifications to G_{max} are explored this way and, if the best-obtained score of the iteration is greater than the best absolute score, then the latter is set to the current S_G and G_{max} is set equal to the current G. The procedure is then repeated from step 6 for the new G_{max}. The algorithm stops when applying all the possible modifications to the DAG G_{max}, obtained in the previous iteration, does not generate an increased S_{max}. In this case, G_{max} constitutes the output of the algorithm.

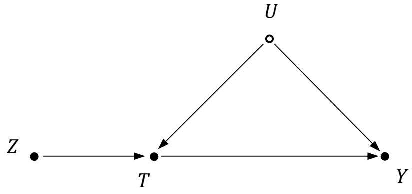
4.4 Causal discovery and Potential Outcomes methods

We have already shown how PO methods can benefit from specifying a causal graph to outline causal relations between variables. If the causal knowledge is available, drawing a causal graph can help assess unconfoundedness in a high dimensional context, using the graphical condition called the back-door criterion. Causal discovery methods constitute an additional resource if we are interested in estimating causal effects with PO methods, but the knowledge of the causal graph is partial or absent. Suppose we want to estimate the effect of treatment T on an outcome Y given a dataset $D(\mathbf{X}, T, Y)$, where **X** are additional random variables, that could directly or indirectly affect T and Y. In addition, let us assume that the available subject matter knowledge concerning the variable causal structure is very limited and thus does not allow drawing a causal graph. In order to estimate causal effects with a PO method such as matching, we have first to assess if unconfoundedness holds. However, since the causal graph over $\{\mathbf{X}, T, Y\}$ is unknown, we cannot directly select an adjustment set **S** that satisfies the back-door criterion.

Causal discovery provides a solution to this scenario. If we cannot exclude the absence of unobserved common causes, we can learn the graph from $D(\mathbf{X}, T, Y)$ employing an algorithm that only requires the faithfulness assumption, such as the FCI algorithm. The algorithm's output can be then used to assess which PO identification strategy is adequate to estimate the causal effect of T on Y. If, instead, it is reasonable to assume both causal sufficiency and faithfulness, we can opt for an algorithm such as the greedy search or PC-stable. In both cases, we know that, if the assumptions hold, the obtained causal structures are asymptotically correct, and a sufficient adjustment set can be selected by applying the back-door criterion. The adjustment set can then be used to derive the interventional distribution through the adjustment formula, or directly estimate the ATE with a method of choice, such as regression, matching or inverse probability weighting.

Alternatively, learning the graph from data could reveal or confirm if a specific PO identification strategy is feasible. Assume that applying a structural learning algorithm on a given dataset generates the DAG in Figure 4.

Figure 4. Instrumental variable DAG



If we are interested in the effect of T on Y, we cannot directly estimate causal effects because of the presence of the unobserved confounder U, and no observed adjustment set that satisfies the back-door criterion. However, the graph configuration reveals that variable Z satisfies the exclusion restriction assumption of instrumental variables described in Section 2. This means that we can employ an IV strategy to achieve causal effect identification. Also in this case, the assumptions required by PO methods are made transparent by causal graph implementation. In this particular example, those assumptions are also strengthened by the structural learning procedure that allows exclusion restrictions to be derived directly from the data.

5 Discussion

Estimating causal effects is a central subject for biomedical and social sciences. However, investigating causal claims is an ambitious objective, especially when dealing with observational data. The most affirmed causality frameworks are Potential Outcomes and Causal Graphs. The two approaches are often contrasted to evaluate which one is most effective. PO methods offer efficient ad hoc solutions to specific causal problems. However, their assumptions are considered difficult to assess, especially as the number of variables increases. On the other hand, causal graphs allow the formalization of complex causal problems in a generalized way. Nevertheless, their high generality can sometimes be perceived as a distance from real empirical problems and incapacity of including context-specific restrictions in the model.

This paper described how the two frameworks could be implemented together in an integrated approach. Causal graphs can be used as a guide to evaluating which PO method can be implemented and if its assumptions hold. The graph can be outlined directly if the causal structure is entirely known or learned from data if the causal knowledge is partial or absent. This versatility guarantees coverage of most empirical problems. The results of PO methods are thus strengthened by causal graphs, since assumptions such as unconfoundedness and exclusion restrictions can be directly assessed from the structure of the DAG. At the same time causal graphs can benefit from all the context-specific identification strategies provided by the literature of Potential Outcomes. Combining the two methodologies thus results in an effective synergic approach that enhances both frameworks' peculiar characteristics.

References

Angrist J. (1990), Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records, *The American Economic Review*, 80, 313-336.

Angrist J. and Pischke J. (2008), *Mostly Harmless Econometrics*. Princeton University Press, Princeton.

Athey S. and Imbens G. (2017), The State of Applied Econometrics: Causality and Policy Evaluation, *Journal Of Economic Perspectives*, 31, 3-32.

Bareinboim E. and Pearl J. (2016), Causal Inference and the Data-Fusion Problem, *Proceedings Of The National Academy Of Sciences*, 113, 7345-7352.

Cinelli C., Forney A. and Pearl J. (2020), A Crash Course in Good and Bad Controls, available at SSRN: https://ssrn.com/abstract=3689437.

Colombo D. and Maathuis M. (2014), Order-Independent Constraint-Based Causal Structure Learning, *Journal of Machine Learning Research*, 15, 3741-3782.

Cox D. (1958), *Planning of Experiments*. Wiley, New York.

Cunningham S. (2021), *Causal Inference: The Mixtape*. Yale University Press, New Haven.

Fisher R. (1949), *The Design of Experiments*. Oliver & Boyd, Edinburgh.

Gasse M., Aussem A. and Elghazel H. (2014), A Hybrid Algorithm for Bayesian Network Structure Learning with Application to Multi-Label Learning, *Expert Systems With Applications*, 41, 6755-6772.

Glymour C., Spirtes P. and Scheines R. (1991), Causal Inference, *Erkenntnis*, 35, 151-189.

Haavelmo T. (1943), The Statistical Implications of a System of Simultaneous Equations, *Econometrica, Journal Of The Econometric Society*, 11, 1-12.

Holland P. (1986), Statistics and Causal Inference, *Journal Of The American Statistical Association*, 81, 945-960.

Huang Y. and Valtorta M. (2006), Pearl's Calculus of Intervention Is Complete, *Proceedings Of The 22nd Conference On Uncertainty In Artificial Intelligence*, 217-224.

Hnermund P. and Bareinboim E. (2019), Causal Inference and Data Fusion in Econometrics, *ArXiv Preprint ArXiv:1912.09104*.

Imbens G. and Rubin D. (2015), *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press, Cambridge.

Imbens G. (2020), Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics, *Journal Of Economic Literature*, 58, 1129-1179.

Margaritis D. (2003), *Learning Bayesian Network Model Structure from Data*. Carnegie-Mellon University, Pittsburgh.

Morgan S. and Winship C. (2015), *Counterfactuals and Causal Inference*. Cambridge University Press, Cambridge.

Neyman J. and Iwaszkiewicz K. (1935), Statistical Problems in Agricultural Experimentation, *Supplement To The Journal Of The Royal Statistical Society*, 2, 107-180.

Pearl J. (2000), *Models, Reasoning and Inference*. Cambridge University Press, Cambridge.

Rosenbaum P. (2018), *Observation and Experiment*. Harvard University Press, Cambridge.

Rubin D. (1974), Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies, *Journal Of Educational Psychology*, 66, 688.

Russell S. and Norvig P. (2009), *Artificial Intelligence: A Modern Approach*. Prentice Hall, Hoboken.

Spirtes P., Glymour C., Scheines R. and Heckerman D. (2000), *Causation, Prediction, and Search*. MIT press, Cambridge.

Splawa-Neyman J., Dabrowska D. and Speed T. (1990), On the Application of Probability Theory to Agricultural Experiments, Essay on Principles, Section 9, *Statistical Science*, 5, 465-472.

Tian J. and Pearl J. (2002), General Identification Condition for Causal Effects, *Aaai/Iaai*, 18, 567-573.

Tinbergen J. (1930), Determination and Interpretation of Supply Curves: An Example, *Zeitschrift Fur Nationalokonomie*, 1, 669-679.

Tsamardinos I., Brown L. and Aliferis C. (2006), The Max-Min Hill-Climbing Bayesian Network Structure Learning Algorithm, *Machine Learning*, 65, 31-78.

Verma T. and Pearl J. (1990), Causal Networks: Semantics and Expressiveness, *Machine Intelligence And Pattern Recognition*, 9, 69-76.

Wright S. Systems of Mating. I. (1921), The Biometric Relations between Parent and Offspring, *Genetics*, 6, 111.

SAPIENZA
UNIVERSITÀ EDITRICE

ANNALI DEL DIPARTIMENTO DI METODI E MODELLI
PER L'ECONOMIA IL TERRITORIO E LA FINANZA
2022 | pp. 45-59

E-ISSN 2611-6634 | ISSN 2385-0825
DOI: 10.13133/2611-6634/1414

Research article

Daniele Mancinelli[*]

MANAGING CASH-IN RISK WITHIN PORTFOLIO INSURANCE STRATEGIES: A REVIEW

Abstract

A *portfolio insurance strategy* is a dynamic hedging process aiming to limit downside risk during a market downturn and allow investors to obtain equity market participation in the upside market. The biggest potential risk of implementing a portfolio insurance strategy is the so-called *cash-in risk*, i.e., the risk that the underlying asset registers huge drops before the portfolio can be rebalanced. In such cases, the value of the insured portfolio would fall below the floor (the insured capital), and the consequence is that the portfolio is fully monetized, not allowing the investor to recover the capital initially invested. First, this paper reviews the main properties of the most used allocation algorithm, the so-called *Constant Proportion Portfolio Insurance* (CPPI), and how the cash-in risk affecting this kind of allocation strategy can be modelled and hedged. Secondly, it describes the main extensions of CPPI proposed in the literature to improve its capability to reduce cash-in risk.

Keywords: OBPI, cash-in risk, CPPI, VPPI, EPPI

---

[*] (Corresponding author) Department of Methods and Models for Territory, Economics and Finance, Sapienza University of Rome, Italy

1 Introduction

Portfolio insurance strategies were first introduced by Rubinstein and Leland (1976) after the collapse of stock markets (the *New York Stock Exchanges Dow Jones Industrial Average* and the *London Stock Exchanges FT 30*) which implied the pension funds withdrawal. In particular, the authors noted ex-post that the presence of an insurance could have convinced the investors not to leave the market, guarantying them the opportunity to take advantage of the rise of the same, an event that happened just a couple of years later. Perold and Sharpe (1988) state that portfolio insurance strategies can be classified into three categories: *option-based strategies*, *option-duplicating strategies* and *derivative-independent strategies*. The primary approach related to the first class of strategies is the so-called *option-Based Portfolio Insurance* (OBPI), which consists of buying a zero-coupon bond with maturity equal to the investment time horizon plus an option written on a risky asset. An option-duplicating strategy is an approach where the option is replicated with a self-financing strategy in order to overcome the lack of liquid options for long maturities. However, the low-interest rate levels which have characterized the markets in recent years are reducing the available risk budgets significantly. Such a market environment forces practitioners to rethink how to build their portfolios to simultaneously offer sustainable equity market participation and capital protection for the initial investment. In this direction, one choice is to consider dynamic risk management tools to protect portions of the initial investment by dynamically allocating wealth into risky and riskless assets. In this framework the *Constant Portfolio Portfolio Insurance* (CPPI) is one of the most used approaches. The CPPI method is obtained by rebalancing an initial portfolio at each observation time, evaluating the present value of the capital to be protected and then investing the available risk budget into risky assets while investing the remaining part of the portfolio in risk-free assets. Despite a significant simplicity and a remarkable ease of implementation, the CPPI strategy suffers a fundamental drawback represented by the risk that, after a severe market draw-down, the risk budget erases. This event is the so-called *cash-in risk* and it has mainly two consequences. Since after the cash-in event, the remaining portfolio is entirely invested into the riskless asset, the CPPI strategy might not guarantee *(i)* the capital initially invested at maturity and, *(i)* an equity market participation in case of subsequent rises of financial market. The paper is structured as follows: in Section 2, we recall the main concepts related to OBPI strategy; in Section 3, we provide an in-depth analysis of CPPI strategy by reviewing how cash-in risk can be modeled and hedged; in Section 4, we describe the main extensions of the CPPI strategy, i.e. the *Time Invariant Portfolio Protection* (TIPP) and the *Variable Proportion Portfolio Insurance* (VPPI), whose aim is to reduce the probability of cash-in event; in Section 5 we focus on an alternative way, more recently introduced, to hedge cash-in risk, based on the use of a particular kind of options; Section 6 concludes.

2 Option based Porfolio Insurance strategies

Option-based portfolio insurance (OBPI) is a methodology characterized by ensuring a minimal terminal portfolio value. Along the lines of Bertrand and Prigent (2005), it is possible to define the

OBPI portfolio process $V^{OBPI} = \left\{V_t^{OBPI}\right\}_{t\in[0,T]}$, with initial value V_0^{OBPI}, as follows:

$$V_t^{OBPI} = qB_t + pC(t, S_t, K), \tag{1}$$

where $q \geq 0$ represents the number of the riskless asset purchased by the investor to protect the capital initially invested, $C(t, S_t, K)$ is the price of the call option, written on the risky asset S_t, having strike price K and maturity T and $p \geq 0$ is the number of call option that can be acquired at time $t = 0$, given the risk budget. The strategy is relatively simple to implement since it is static, i.e. no trading occurs in $[0, T]$. From eq. (1) it is straightforward to show that K represents the wealth the investor wishes to recover at maturity T. However, it may happen that European options, whose strike price is equal to the amount of wealth to be immunized, are not traded on the market. This implies that the investor must synthetically replicate the payoff at maturity of the option using a hedging strategy. Perfect hedging can be achieved only by assuming that the market is complete. However, it is well known that various sources of market incompleteness exist in terms of stochastic volatility and trading restrictions, making the contingent claim payoff unattainable. This implies that the standard OBPI approach is not always viable. This issue explains why dynamic Portfolio Insurance strategies such as CPPI, which will be examined in 3, have become prominent among practitioners.

3 The Constant Proportion Portfolio Insurance

The objective of the Constant Proportion Portfolio Insurance (CPPI) investors is twofold: participating in the upside potential of the risky reference portfolio, e.g. a market index, and at the same time, ensuring that the value of the portfolio at maturity V_T^{CPPI} is higher or equal to a guaranteed amount G ($V_{CPPI}^T \geq G$). The guarantee G is a proportion $PL \in (0, 1]$ of the initially invested amount V_0^{CPPI}. These two goals are realized by dynamically allocating the initial wealth V_0^{CPPI} between a risk-free asset and a market index. In order to define the CPPI portfolio process, we begin by specifying the so-called *floor* F_t, representing the lowest acceptable value for the portfolio for each instant of time $t \in [0, t]$. The floor F_t is given by

$$F_t = F_T e^{-r(T-t)}, \tag{2}$$

where $F_T = G = PL \cdot V_0^{CPPI}$ and r is the risk-free interest rate. The next step is to compute the *cushion* C_t, which is the difference between the portfolio value V_t and the floor F_t. The exposure to the risky asset E_t is given by the product between the cushion C_t and the multiplier $m \in \mathbb{R}^+$. The latter parameter amplifies the risk budget and is exogenously set by the investor at $t = 0$. Since the strategy is self-financing, the remaining part of the portfolio, i.e. $V_t^{CPPI} - E_t$, is invested into the risk-free asset. Then, the proportion of wealth invested to the risky asset for each instant of time $t \in [0, T]$ can be written as:

$$\alpha_t^{CPPI} = \min\left\{\frac{m(V_t^{CPPI} - F_t)^+}{V_t^{CPPI}}, LEV\right\}, \tag{3}$$

$$\beta_t^{CPPI} = 1 - \alpha_t^{CPPI}. \tag{4}$$

where $LEV \in (0, 2]$ is the *maximum leverage factor*: to avoid excessive equity exposure, E_t is bounded to be at most $LEV \cdot V_t^{CPPI}$. We start by considering the case in which the CPPI portfolio is continuously rebalanced, meaning that the Exposure E_t and the investment in the riskless asset B_t are continuously adjusted. The main properties and the structure of the continuous-time CPPI allocation strategy, summarized 3.1, have been extensively studied in Black and Perold (1992).

3.1 CPPI with continuous rebalancing

In order to define the continuous-time CPPI portfolio process, we begin by specifing the so-called floor process $F = \{F_t\}_{t\in[0,T]}$ whose dynamic is given by

$$dF_t = rF_t dt, \tag{5}$$

with initial value $F_0 = V_0^{CPPI} \cdot PL \cdot e^{-rT}$. Then we define the process $V^{CPPI} = \left\{V_t^{CPPI}\right\}_{t\in[0,T]}$ with initial value V_0^{CPPI}, representing the portfolio value associated to CPPI strategy, namely:

$$V_t^{CPPI} = \alpha_t^{CPPI} S_t + \beta_t^{CPPI} B_t, \tag{6}$$

where α_t^{CPPI} (resp. β_t^{CPPI}) is depicted in eq. (3) (resp. eq. (4)). Furthermore, since the CPPI strategy is self-financing, the dynamics of V_t^{CPPI} is given by

$$dV_t^{CPPI} = \alpha_t^{CPPI} dS_t + \beta_t^{CPPI} dB_t. \tag{7}$$

Moreover, if we assume that the risky asset S_t follows a geometric Brownian motion, i.e.:

$$dS_t = \mu S_t dt + \sigma S_t dW_t^{\mathbb{P}}, \tag{8}$$

where $W = \{W_t\}_{t\in[0,T]}$ is a standard Brownian motion with respect to the real world measure $\mathbb{P}$, $\mu \in \mathbb{R}$ s.t. $\mu > r \geq 0$ and $\sigma \in \mathbb{R}_+$, Black and Perold (1992) explicitly derived the SDE satisfied by the cushion process given by $C = \{C_t\}_{t\in[0,T]}$, given by

$$\frac{dC_t}{C_t} = \bigg(r + m(\mu - r)\bigg)dt + m\sigma dW_t^{\mathbb{P}}. \tag{9}$$

From eq. (9), it is straightforward to show that the CPPI portfolio value is:

$$V_t^{CPPI} = (V_0 - G \cdot e^{-rT})e^{[r - m(r - \frac{\sigma^2}{2}) - \frac{m^2\sigma^2}{2}]t}\left(\frac{S_t}{S_0}\right)^m + G \cdot e^{-rT},\ t \in [0, T]. \tag{10}$$

Eq. (10) illustrates that within this framework, CPPI portfolio is equivalent to taking a long position in a zero-coupon bond with nominal value G in order to guarantee the capital at maturity, and investing the remaining part into a risky asset which has m times the excess return of S and is perfectly correlated with S. Moreover, it shows that the portfolio protection is efficient almost surely: the terminal value of the CPPI strategy is higher than the guarantee with probability one, regardeless of multiplier value. Indeed, the expected value of a CPPI-insured portfolio at maturity is equal to:

$$\mathbb{E}[V_T] = G + (V_0 - G \cdot e^{-rT})e^{[r+m(\mu-r)]T}. \tag{11}$$

which is always greater or equal with respect to the amount of capital that the investor wishes to recover at the end of the investment time horizon.

As highlighted in Balder et al. (2009); Cont and Tankov (2009), CPPI managers widely recognize the possibility of reaching the floor: there is a non-zero probability that, during a sudden downside movement of the underlying asset, the fund manager will not have time to readjust the portfolio, which crashes through the floor. This implies that the remaining portfolio value will be shifted entirely to the risk-free asset. Hence, it is no longer ensured that the strategy outperforms the prescribed floor. As mentioned in Section 1, this risk is known as cash-in risk.

Measuring the risk that the CPPI strategy is less than the floor is of practical importance for at least two reasons. Firstly, a CPPI strategy is combined with a guarantee for the investor: even if the floor has been broken, the CPPI issuer must pay the guaranteed amount F_T. Since the CPPI has fallen below the floor during the investment period, i.e. $V_T < F_T$, the issuer has to pay out more than the CPPI is worth. For this reason, an additional option can be added. Such an option is exercised if the value of the CPPI is below the floor. Secondly, CPPI strategies can be used to protect return guarantees embedded in unit-linked life insurance contracts. In the above case, maturity can be interpreted as retirement age, and guarantee as the amount which is at least needed by the insured.

The formal proof that there exist only two sources of cash-in risk is given in Schied (2013). The first source, extensively studied in Balder et al. (2009), is represented by discrete rebalancing of the CPPI portfolio. The second source of cash-in risk, modeled for the first time by Cont and Tankov (2009), is given from the fact that the price of the underlying risky asset may experience downward jumps.

3.2 CPPI with discrete time rebalancing

Let τ denote a sequence of equally spaced instants of time belonging to the interval $[0, T]$, i.e.:

$$\tau = \{t_0 = 0 < t_1 < \cdots < t_{n-1} = T\}, \tag{12}$$

where $t_{k+1} - t_k = \frac{T}{n}$ for $k = 0, \ldots, n-1$. We impose the rescrition that trading is only possible immediately after $t_k \in \tau$. This implies that the number of shares held in the risky asset is constant over the interval $(t_k, t_{k+1}]$ for $k = 0, \ldots, n-1$. However, the portion of CPPI portfolio invested in risky asset changes as risky asset price fluctuates. Thus, it is necessary to consider the number of shares held in the risky asset $\phi^{(S)}$ and the number of risk-free bond $\phi^{(B)}$. Along the lines of Balder et al. (2009), we indicate by $\phi_\tau = (\phi^{(S),\tau}, \phi^{(B),\tau})$ a discrete time CPPI if, for $t \in (t_k, t_{k+1}]$ and $k = 0, \ldots, n-1$, $\phi_t^{(S),\tau} := \max\left\{\frac{mC^\tau_{t_k}}{S_{t_k}}, 0\right\}$, where the cushion is given by

$$C^\tau_{t_{k+1}} = C^\tau_{t_0} \prod_{i=1}^{\min\{\nu,k+1\}} \left(m\frac{S_{t_i}}{S_{t_{i-1}}} - (m-1) \right), \tag{13}$$

with

$$\nu := \min\left\{t_k \in \tau | V^{\tau}_{t_k} - G \leq 0\right\}, \tag{14}$$

and $\nu = \infty$ if the minimum is not attained. Within this framework, the authors quantify the cash-in risk by computing the following quantities:

1. the *local shortfall probability*,

$$\mathbb{P}^{LSF} := \mathbb{P}(V^{\tau}_{t_{k+1}} \leq F_{t_{k+1}} | V^{\tau}_{t_k} > F_{t_k}) = \mathcal{N}(-d_2),$$

   where $d_2 = \frac{\ln \frac{m}{m-1} + (\mu - r)\frac{T}{n} - \frac{\sigma^2}{2}\frac{T}{n}}{\sigma\sqrt{\frac{T}{n}}}$,

2. the *shortfall probability*,

$$\mathbb{P}^{SF} := 1 - (1 - \mathbb{P}^{LSF})^n,$$

3. the *expected shortfall*,

$$ES := \mathbb{E}[G - V^{\tau}_T | V^{\tau}_T \leq G] = G + (V_0 - F_0)\left[E^n_1 + e^{-r\frac{T}{n}} E_2 \frac{e^{rT} - E^n_1}{1 - E_1 e^{-r\frac{T}{n}}}\right],$$

   where

$$\begin{aligned} E_1 &:= m e^{\mu\frac{T}{n}} \mathcal{N}(d_1) - e^{r\frac{T}{n}}(m-1)\mathcal{N}(d_2), \\ E_2 &:= e^{r\frac{T}{n}}[1 + m(e^{(\mu - r)\frac{T}{n}} - 1)] - E_1. \end{aligned}$$

   with $d_1 := d_2 + \sigma\sqrt{\frac{T}{n}}$.

Within this framework, two solutions have been proposed to ensure the effectiveness of the CPPI strategy. The first one is the following: given an estimate for μ and σ, it is possible to determine the value of the multiplier m and the number of rebalances n of $\phi^{(S),\tau}$ and $\phi^{(B),\tau}$ over $[0, T]$ such that the probability of falling below the guarantee G is bounded above a confidence level γ. The second one has been proposed by Bertrand and Prigent (2016). It is based on large deviation methods that the authors use to estimate the possible losses between two consecutive trading dates.

3.3 CPPI in presence of jumps in asset prices

As mentioned in Section 3.1, the second alternative to model cash-in risk is to allow for jumps in the risky asset dynamics without relaxing the continuous trading assumptions. CPPI strategies with the presence of jumps in stock prices were considered for the first time by Prigent and Tahar (2005) in a jump-diffusion model with finite intensity activity. However, the approach by Prigent and Tahar (2005) is to consider as a slight modification of the CPPI strategy, which incorporates an additional guarantee whenever the portfolio falls below the floor.

The first study devoted to quantifying cash-in risk within CPPI strategy by considering a more general framework including infinite activity jumps and stochastic volatility, is given by Cont and Tankov (2009). The reason behind introducing the above models is to highlight that cash-in risk cannot be attributed exclusively to trading restriction. This could give the wrong impression that this risk can be eliminated by more frequent rebalancing. Indeed, Cont and Tankov (2009) argues that by considering jumps in the risky asset price dynamic, there is a non-negligible residual cash-in risk for CPPI, even in continuous trading.
As in Section 3.1, assume a continuous time market model. Hence, we have to consider a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = \{\mathcal{F}_t\}_{0 \leq t \leq T}, \mathbb{P})$ with two $\mathbb{F}$-adapted processes describing the evolution for the riskless asset B_t, and the risky asset S_t. As in the previous Sections, we assume that r is constant over the investment time horizon $[0,T]$. Then, B_t has the following deterministic evolution:

$$dB_t = rB_t dt, \ t \in [0,T], \tag{15}$$

with $B_0 = b$. In this case we assume that the price process for the risky asset S_t is

$$dS_t = S_{t-} dZ_t, \tag{16}$$

where Z_t is a possible discontinous driving process, modeled as semimartingale. Moreover, in order to ensure the positivity of the price process, we assume that $\Delta Z_t > -1$. In a continuous-time setting the stopping time defined in eq. (14) becomes

$$\nu = \inf \{t \geq 0, V_t \leq F_t\}. \tag{17}$$

If $t < \nu$, the CPPI strategy portfolio value satisfies

$$dV_t = m(V_{t-} - F_t)\frac{dS_t}{S_{t-}} + \Big(V_{t-} - m(V_{t-} - F_t)\Big)\frac{dB_t}{B_t}, \tag{18}$$

which implies the following dynamic for the cushion

$$\frac{dC_t}{C_{t-}} = m dZ_t + (1-m) r dt. \tag{19}$$

Introducing the discounted cushion $C_t^* = \frac{C_t}{B_t}$ and applying Itô formula, we find

$$\frac{dC_t^*}{C_{t-}^*} = m(dZ_t - r dt). \tag{20}$$

Defining $L_t = L_0 + \int_0^t dZ_s - rt$, eq. (20) can be rewritten as $\frac{dC_t^*}{C_{t-}^*} = m dL_t$ or, equivalently, $C_t^* = C_0 \mathcal{E}(mL)_t$ where $\mathcal{E}$ denotes the stochastic (Dolans-Dade) exponential defined by

$$\frac{d\mathcal{E}(mL)_t}{\mathcal{E}(mL)_{t-}} = m dL_t. \tag{21}$$

If $\nu \geq t$, according to the definition of CPPI strategy, the value of the portfolio is entirely invested into the risk-free asset, in order to prevent further losses. This means that, when $\nu \geq t$, the value of C_t^* remains constant. Then, for any $t \in [0, T]$ we introduce a new process $\tilde{C}_t$ defined as the *stopped* process of C^* such that $\tilde{C}_t = C^*_{t\wedge\nu}$, where $t \wedge \nu := \min\{t, \nu\}$ and it can be explicitly written as:

$$\tilde{C}_t = C_0^* \mathcal{E}(mL)_{t\wedge\nu}. \tag{22}$$

Hence, the CPPI portfolio may fall below the floor, even if the exposure and riskless asset investment are adjusted continuously. This is due to the fact that the stochastic exponential in eq. (22) can become negative in the presence of negative jumps of sufficient size of stock price.Within this framework, Cont and Tankov (2009) quantify in a meaningful way the cash-in risk by establishing a direct relationship between the value of the multiplier m and the risk of the insured portfolio. This allows choosing the multiplier based on the risk tolerance of the investor. In particular, the authors provide a Fourier transform method to compute the losses distribution and various risk measures (e.g., Value-at-Risk, expected loss, or the probability of loss) over a given time horizon. Moreover, they extend the framework described in this Section by adding stochastic volatility.

4 Some extensions of CPPI allocation strategy

In order to mitigate the cash-in risk which affects the CPPI strategy, several solutions have been proposed. In particular, the financial literature derived mainly two extensions of the CPPI allocation algorithm. The first modification, proposed by Estep and Kritzman (1988), is the so-called Time-Invariant Portfolio Protection (TIPP). It is based on the same rules as the CPPI allocation algorithm except for the floor computation. In fact, in this case, it is no longer linked to the risk-free interest rate r but to the portfolio's value. The second modification, proposed by Lee et al. (2008), is the so-called Variable Proportion Portfolio Insurance (VPPI), which concerns how the multiplier is set. In the standard CPPI, the multiplier m is fixed at $t = 0$, based on the investor's risk aversion, and remains constant throughout the investment time horizon $[0, T]$. Instead, the VPPI allows the multiplier to vary over time according to specific factors, such as the volatility of the risky asset underlying the strategy. In what follows, we will discuss how these new strategies affect the amount invested into risky assets over time and, consequently, their ability to guarantee at least G at maturity and equity market participation in the event of bull markets.

4.1 Time Invariant Portfolio Protection (TIPP)

Standard results about CPPI strategies are based on the assumption that the floor F_t evolves like the riskless asset B_t. However, this assumption is quite stringent since it makes the CPPI performances highly path dependent; any gains at a particular time $t \in [0, T]$ can be lost if the underlying asset price registers a considerable fall. To address this issue, Estep and Kritzman (1988) proposed the Time-Invariant Portfolio Protection (TIPP), i.e. a modified version of CPPI with a ratchet mechanism, able to lock in a proportion of the upside performance. In particular, this mechanism consists of a stochastic time-varying definition of the floor. The TIPP floor F_t is defined as the maximum between the usual CPPI floor and a percentage of the historical maximum portfolio value. The new floor $\tilde{F}_t$ satisfies

$$\tilde{F}_t = \max\left\{F_t, PL \cdot \sup_{s\leq t} V_s\right\}, \ \forall t \in [0, T], \tag{23}$$

where PL is the protection level. In this way, the floor jumps up with the portfolio value to reduce risky asset allocation when the market peaks. This new floor adjustment has some consequences on the allocation of the risky asset over time, especially in market scenarios where the risk-free interest rates are very low and sudden rises and falls of the risky asset might occur. Such a background sheds light on one of the main issues related to using CPPI: when the value of the risky asset increases, the CPPI portfolio value increases accordingly. If the level of the risk-free interest rate is low, then the growth rate of the CPPI portfolio will be higher than the corresponding growth rate of the floor. This implies that there is no longer significant portfolio protection after a very short period. Such a drawback is overcome by the TIPP allocation strategy, thanks to dynamics in eq. (23). Indeed, the growth rate of the TIPP floor is comparable to the portfolio for every $t \in [0, T]$, even if the market growth is sustained. Consequently, the TIPP exposure to the risky asset will generally be lower than the CPPI one. It will change more smoothly over time, furnishing better downside protection to the investor. However, in case of favourable market conditions, the TIPP strategy's overall return will be generally lower w.r.t. the standard CPPI one. This is in line with the empirical analysis carried out by Dichtl et al. (2017), in which the authors conclude that TIPP cannot be seen as an improved CPPI. Indeed, the former exhibits a significantly inferior performance for all applied performance measures, even if it can furnish better protection against cash-in risk.

4.2 Variable Proportion Portfolio Insurance (VPPI)

Instead of continuously rebalancing the floor, it is reasonable to make the multiplier dynamic. This change makes the CPPI a Variable Proportion Portfolio Insurance, the so-called VPPI. One of the main reasons to consider such an extension is to allow the investor to better adapt his portfolio strategy to market fluctuations. For example, suppose that for the first year of a five-year global management period, the forecast for the stock index's performance is that significant sudden drops can occur. In this case, at $t = 0$, the investor has to choose a relatively low multiplier. However, suppose substantial rises will occur in the future. In that case, the exposure corresponding to a small value of the multiplier initially chosen will not provide the opportunity to benefit from a bullish market. On the contrary, if the initial value of the multiplier is relatively high, sudden significant drops will imply that the portfolio may break the floor. This will imply that, at maturity, the investor might only recover G. The possibility of reducing the multiple during the investment period can prevent such an unfavourable event. This extension of the classical CPPI allocation strategy was introduced by Lee et al. (2008). In order to keep the simplicity of the model, the authors proposed a particular kind of VPPI strategy, the *Exponential Proportion Portfolio Insurance* (EPPI), where the multiplier changes according to the following criteria. In $t = 0$, the investor has to fix a reference price for the risky asset underlying the strategy. Such a reference price is the value of the risky asset before the portfolio rebalance, $S^{(0)}$. If the value of the risky asset after the portfolio rebalancing,

$S^{(1)}$ is different from $S^{(0)}$, the multiplier changes according to the following rule:

$$m_t = \eta + \exp\left\{a \ln\left(\frac{S^{(1)}}{S^{(0)}}\right)\right\},\ t \in [0, T], \tag{24}$$

where $\eta > 1$ is an arbitrary constant, and $e^{a \ln \frac{S^{(1)}}{S^{(0)}}}$ is the so called *dynamic multiple adjustment factor* (DMAF), with $a > 1$. The parameter a acts as the magnifier of the DMAF. More precisely, it is set greater than 1 for the enlargement effect on the number of holding shares when the stock price goes up, and the shrinkage effect on the number of holding shares when the stock price goes down w.r.t. the reference stock price $S^{(0)}$. Thus, including a DMAF into the multiplier could create a higher convex nature of the strategy, i.e. when the stock price goes up, the mechanism of the EPPI strategy creates more holding shares to perform an upside capture. In contrast, when the stock price goes down, the strategy creates fewer holding shares to provide a downside protection. However, the EPPI strategy is not the only extension of the CPPI, which allows for a dynamic multiplier. Indeed, a wide range of models has been proposed directly linked to a risk management approach. In particular, the baseline assumption of these kinds of models is to fix the multiplier at each rebalancing date by considering a *local quantile guarantee condition* posed by

$$\mathbb{P}_{t_k}(C_{t_{k+1}} > 0 | C_{t_k} > 0) \geq 1 - \varepsilon, \tag{25}$$

where $\mathbb{P}_{t_k}(\cdot)$ denotes the conditional probability with respect to $\mathcal{F}_{t_k}$ and the parameter ε denotes an exogenously specified upper bound on the local shortfall probability. We further observe that

$$\mathbb{P}_{t_k}(C_{t_{k+1}} > 0 | C_{t_k} > 0) = \mathbb{P}_{t_k}\left(m_{t_k}\frac{S_{t_{k+1}}}{S_{t_k}} - (m_{t_k} - 1) > 0\right). \tag{26}$$

Eq.(26) implies an upper bound $\bar{m}_{t_k}$ on the admissible multiplier, i.e. a gap control affords to limit the multiplier at t_k by $\bar{m}_{t_k}$. Denoting by $\tilde{F}_{t_k}$ the marginal distribution function of the standardized simple return $\frac{S_{t_{k+1}}}{S_{t_k}} - 1$, the upper bound is given by

$$\bar{m}_{t_k} = \left| E_{t_k}\left[\frac{S_{t_{k+1}}}{S_{t_k}}\right] - 1 + \sqrt{Var_{t_k}\left[\frac{S_{t_{k+1}}}{S_{t_k}}\right]}\tilde{F}_{t_k}^{-1}(\varepsilon)\right|^{-1}. \tag{27}$$

The condition in eq. (27) is used in Hamidi et al. (2009), who estimate the conditional upper multiplier by means of the *Value at Risk* (VaR). In particular, they resort to eight different methods of VaR calculations: one non-parametric method using the historical simulation approach; three parametric methods based on distributional assumptions: namely, the normal VaR, the RiskMetrics VaR based on the normal distribution, and the GARCH VaR based on the Student-t distribution; four semi-parametric methods using quantile regression to estimate the conditional autoregressive VaR (CAViaR): namely, the symmetric Absolute Value CAViaR, the Asymmetric Slope CAViaR, the IGARCH(1,1) CAViaR, and the Adaptive CAViaR. Afterwards, Hamidi et al. (2014) proposed a generalization of this class of models in which the conditional multiplier is based on a coherent risk measure, the expected shortfall. In this case, they estimate the conditional upper bound for the multiplier $\bar{m}_{t_k}$ using a Dynamic AutoRegressive Expectile (DARE) model.

5 Hedging cash-in risk through options

Another way to hedge cash-in risk within portfolio insurance strategies is to use options. Let us consider the Vanilla options written on the CPPI portfolio as an example. For instance, taking a long position on an at-the-money put option on the CPPI portfolio with a strike price at least equal to the minimum value that the investor requires at maturity is a natural way to hedge cash-in risk. Similarly, taking a long position on an at-the-money call option on the CPPI portfolio is a natural way to invest in a CPPI portfolio, preserving the capability to not pursue forward the investment in the case of closed out. The first option pricing model for the CPPI strategy was proposed by Escobar et al. (2011) and generalized by Wang and Tsoi (2013).

5.1 Pricing option on CPPI

As described in Section 3.2, a possible way to model cash-in risk is to allow trading of the CPPI portfolio only on specific dates. For this reason, Escobar et al. (2011), to develop a pricing formula for European options written on CPPI, decided to model the risky asset underlying the strategy using a geometric Brownian motion, restricting trading to discrete-time. The discrete-time process describing the evolution of the risky asset under the risk-neutral probability measure $\mathbb{Q}$ is:

$$\frac{S_{t_k}}{S_{t_{k-1}}} = \exp\left\{\left(r - \frac{\sigma}{2}\right)\Delta t + \sigma W_{t_k}\right\}, \tag{28}$$

where $W_{t_k} \sim N(0, \Delta t)$ and τ, as in eq. (12), is a sequence of equally spaced instants of time of the interval $[0, T]$, i.e $\tau = \{t_0 = 0 < t_1 < \cdots < t_{N-1} < T_N = T\}$ and $\Delta t = \frac{T}{n}$ for $k = 1, \ldots, N-1$. Within this framework the value $V^{\tau}_{t_{k+1}}$ of the CPPI portfolio is:

$$V^{\tau}_{t_{k+1}} = e^{r(t_{k+1} - \min\{\nu, k+1\})}\left((V_0 - F_0)\prod_{i=1}^{\min\{\nu,k+1\}}\left(m\frac{S_{t_i}}{S_{t_{i-1}}} - (m-1)e^{r\Delta t}\right) + F_{t_{\min\{\nu,k+1\}}}\right), \tag{29}$$

where ν given in eq. (14) is the first time the portfolio value breaks through the floor. In order to compute the price of the European option, the authors first derived the price of discrete time CPPI by making the following assumption. Since the CPPI strategy is by definition self-financing, then under the risk-neutral probability measure the expected terminal portfolio value is $\mathbb{E}^{\mathbb{Q}}[V_T] = e^{rT}V_0$. However, for the investor, the value of the CPPI strategy is not equal to V_0. As pointed out in Section 3.1, the CPPI is combined with a guarantee for the investor; this means that the issuer of the CPPI should guarantee a payoff equal to or greater than the floor and therefore has to carry the cash-in risk. Consequently, the payoff at maturity of the discrete CPPI is:

$$CPPI_T = \max\{V_T, F_T\}. \tag{30}$$

Then, the expected value at maturity T, under the risk-neutral probability measure $\mathbb{Q}$ can be written as:

$$\mathbb{E}^{\mathbb{Q}}[CPPI] = \mathbb{E}^{\mathbb{Q}}[CPPI_T|C_1] + \mathbb{E}^{\mathbb{Q}}[CPPI_T|C_2], \tag{31}$$

where $\mathbb{E}^{\mathbb{Q}}[CPPI_T|C_1]$ (resp. $\mathbb{E}^{\mathbb{Q}}[CPPI_T|C_2]$) is the expected value of the portfolio at maturity when the portfolio does not fall below the floor (resp. the expected value of the portfolio when the CPPI has fallen below the floor) over $[0, T]$. When the underlying risky asset follows the process depicted in eq. (28), both of them can be evaluated in closed form. Indeed, the authors proved that:

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}[CPPI_T|C_1] &= F_0\mathcal{N}(s_2)^N + (V_0 - F_0)[m\mathcal{N}(s_1) - (m-1)\mathcal{N}(s_2)]^N, & (32)\\ \mathbb{E}^{\mathbb{Q}}[CPPI_T|C_2] &= F_0(1 - \mathcal{N}(s_2)^N), & (33) \end{aligned}$$

where $s_1 = \frac{\ln m - \ln(m-1) + \frac{\sigma^2}{2}\Delta t}{\sigma\sqrt{\Delta t}}$ and $s_2 = s_1 - \sigma\Delta t$. In order to compute the price of the European option on discrete CPPI, the authors distinguish two cases. The first case is when the strike price, K, is equal to the value of the guarantee at maturity F_T. In this case, the option on CPPI ends up in the money at maturity T if the strategy has not defaulted until maturity, i.e. $C^{\tau}_{t_{k+1}} > 0, k = 1, \ldots, N-1$. Then, a portfolio composed by the option on a CPPI with $K = F_T$ and a zero coupon bond with nominal value K, is exactly equal to a CPPI with floor F. For this reason when $K = F_T$, the price of the call option on CPPI with discrete rebalancing is exacty equal to the difference between the expected value of discrete CPPI in $t = 0$ given by eq. (32) and F_0:

$$C_t = (V_0 - F_0)[m\mathcal{N}(s_1) - (m-1)\mathcal{N}(s_2)]^N. \tag{34}$$

The second case is when the strike price is greater than the guarantee amount at maturity ($K \geq F_T$). In this more general case the option ends up in the money at maturity T if the CPPI has not fallen below the floor in t_k, $k = 1, \ldots, N-1$, and if the value V_T is greater than the strike price K. For this reason the option pricing formula defined in eq. (34) becomes:

$$C_t = e^{-rT}\mathbb{E}^{\mathbb{Q}}\left[\left(F_T - K + (V_0 - F_0)\prod_{i=1}^{N}(me^{(r-\frac{\sigma^2}{2})\Delta t + \sigma W_{t_i}} - (m-1)e^{r\Delta t})\right)^+ \prod_{i=1}^{N-1} 1_{\{V_{t_l} > F_{t_l}\}}\right]. \tag{35}$$

In this more general case, the option price on discrete CPPI and very useful sensitivities like Delta, Gamma, Vega, Rho and Theta can be obtained using Monte Carlo simulations in eq. (35). As explained in Section 3, the second source of cash-in risk for the CPPI strategy can be modelled by adding jumps into the dynamic of the underlying risky asset. Because of this, an alternative way to price the European option on CPPI is to consider a continuous time setting and a jump-diffusion model for the dynamics of the underlying risky asset. Wang and Tsoi (2013) provided the first work in this direction. In particular, they developed a semi-closed formula to price European options written on the CPPI strategy when the underlying risky asset evolves according to the model of Merton (1976). Moreover, since the market is not complete in a jump-diffusion setting, the payoff of these particular contingent claims is no longer attainable. For this reason, the authors developed a particular hedging strategy for this model, the so-called mean-variance hedging.

5.2 Structured product written on a modified version of the CPPI strategy

Using options linked to portfolio insurance strategies can be considered a suitable way to obtain downside protection when the underlying risky asset has experienced heavy losses during the investment time horizon. However, this method cannot offer equity market participation if the risky asset recovers nicely after a severe market draw-down. Di Persio et al. (2021) introduced a modified version of the standard CPPI by defining a minimum threshold always invested in the risky asset to overcome the latter scenario. They called this new kind of strategy CPPI with *guaranteed minimum equity exposure* (GMEE-CPPI). In particular, they introduced the GMEE α_{min} with $0 \leq \alpha_{min} \leq 1$ in eq. (3) obtaining

$$\alpha_t := \max\left\{\min\left\{\frac{m(V_t^{CPPI} - F_t)}{V_t^{CPPI}}, LEV\right\}, \alpha_{min}\right\},\ t \in [0,T]. \tag{36}$$

Thanks to eq. (36), the equity participation will never go below α_{min} even in case of a severe market drop. However, at the same time, it would mean that this adjusted CPPI allocation implemented in a real portfolio might not be able to protect the invested capital. For this reason, the authors introduced a new structured product consisting in a combination of a CPPI strategy and an OBPI one. In particular, this new strategy can be summarized into the following key points:

(i) a significant proportion of the initial portfolio value is invested in time-congruent zero coupon bonds following the classical OBPI approach described in Section 2,

(ii) the remaining part of the portfolio is put into an exotic call option linked to a GMEE-CPPI strategy where the CPPI portfolio has an equity index as a risky asset.

In particular, the authors provided historical simulations showing how the risk-return profile changes according to the market environment and describing the option price behaviours under different frameworks, namely, when the underlying is a pure risky asset, a CPPI strategy, or a CPPIGMEE based one.

The authors argued that this new method provides a valuable compromise between a pure risky asset investment strategy and a traditional CPPI one. Indeed, this innovative method overcame, at the same time, different problems since the use of the OBPI can drastically reduce the cash-in risk, and the use of GMEE-CPPI as the underlying risky asset can ensure some equity market participation.

6 Conclusion and further research

In the present paper, we have reviewed the main properties of Portfolio Insurance strategies and the main risk affecting these particular kinds of dynamic hedging processes, the so-called cash-in risk. We have focused on modelling and hedging cash-in risk within the prominent portfolio insurance strategy, the Constant Proportion Portfolio Insurance (CPPI). First, we have introduced the main properties of the CPPI strategy within three different frameworks. In particular, we consider the CPPI strategy when the underlying risky asset follows *(i)* a geometric Brownian motion, *(ii)* a more general jump-diffusion process, and *(iii)* a geometric Brownian motion where trading is restricted to discrete time. We provide an in-depth analysis of how cash-in risk can be modelled and hedged for each setting mentioned above. Then, we analyzed the most important modifications of the CPPI strategy designed to reduce the probability of cash-in risk. The first is obtained by considering a stochastic time-varying definition of the floor process, the so-called Time-Invariant Portfolio Protection (TIPP). The second one is obtained by varying the multiplier over time according to market fluctuations, the so-called Variable Proportion Portfolio Insurance (VPPI). Lastly, we have reviewed another way to hedge cash-in risk using options linked to the CPPI strategy. Within this framework, we analyze a two-step principal protection strategy obtained by combining a modification of the CPPI, the so-called CPPI with guaranteed minimum equity exposure (GMEE-CPPI), and a classical OBPI strategy. As discussed in Section 5.2, this novel approach, introduced by Di Persio et al. (2021), represents a concrete innovation in the literature related to portfolio protection strategies. Along this line of research, further contributions could be made by including more structured derivatives evaluated concerning general stochastic volatility models with the presence of jumps. Moreover, other inputs for further research can be represented by furnishing a sensitivity analysis of the CPPI-GMEE approach w.r.t. to changes in market parameters. Lastly, another line of research can be represented by comparing options on CPPI with options on other dynamic asset allocation strategies, such as the VolTarget ones, allowing the CPPI-GMEE to have lock-in elements.

References

Balder, S., Brandl, M. and Mahayni, A. (2009), Effectiveness of CPPI strategies under discrete-time trading, *Journal of Economic Dynamics and Control*. 13, 204-220.

Bertrand, P. and Prigent, J.L. (2005), Portfolio insurance strategies: OBPI versus CPPI, *Finance*, 26, 5-32.

Bertrand, P. and Prigent, J.L. (2016), Portfolio insurance: the extreme value approach to the CPPI method, *Extreme Events in Finance: A Handbook of Extreme Value Theory and Its Applications*, 465.

Black, F. and Perold, A.F. (1992), Theory of constant proportion portfolio insurance, *Journal of Economic Dynamics and Control*, 16, 403-462.

Cont, R. and Tankov, P. (2009), Constant Proportion Portfolio Insurance in the presence of jumps in asset prices, *Mathematical Finance*, 19, 379-401.

Dichtl, H., Drobetz, W. and Wambach, M. (2017), A bootstrap-based comparison of portfolio insurance strategies, *The European Journal of Finance*, 23, 31-59.

Di Persio, L., Oliva, I. and Wallbaum, K. (2021), Options on constant proportion portfolio insurance with guaranteed minimum equity exposure, *Applied Stochastic Models in Business and Industry*, 37, 98-112.

Escobar, M., Kiechle, A., Seco, L. and Zagst, R. (2011), Option on a CPPI, *International Mathematical Forum*, 6, 229-262.

Estep, T. and Kritzman, M. (1988), TIPP: Insurance without complexity, *Journal of Portfolio Management*, 14, 38.

Hamidi, B., Bertrand, M. and Prigent, J.L. (2009), A risk management approach for portfolio insurance strategies, *Proceedings of the 1st EIF International Financial Research Forum, Economica*.

Hamidi, B., Bertrand, M. and Prigent, J.L. (2014), A dynamic autoregressive expectile for time-invariant portfolio protection strategies, *Journal of Economic Dynamics and Control*, 46, 1-29.

Lee, H.I., Chiang, M.H. and Hsu, H. (2008), A new choice of dynamic asset management: the variable proportion portfolio insurance, *Applied Economics*, 40, 2135-2146.

Merton, R.C. (1976), Option pricing when underlying stock returns are discontinuous, *Journal of financial economics*, 3, 125-144.

Perold, A.F. and Sharpe, W.F. (1988), Dynamic strategies for asset allocation, *Financial Analysts Journal*, 44, 16-27.

Prigent, J.L. and Tahar, F. (2005), CPPI with cushion insurance, *THEMA University of Cergy-Pontoise working paper*.

Rubinstein, M. and Leland, H.E. (1976), The evolution of portfolio insurance, *Luskin, D.L., Ed., Dynamic Hedging: A Guide to Portfolio Insurance*.

Schied, A. (2013), Model-free CPPI, *Journal of Economic Dynamics and Control*, 40, 84-94.

Wang, M. and Tsoi, A. (2013), CPPI in the jump-diffusion model, *State-Space Models*.

SAPIENZA
UNIVERSITÀ EDITRICE

ANNALI DEL DIPARTIMENTO DI METODI E MODELLI
PER L'ECONOMIA IL TERRITORIO E LA FINANZA
2022 | pp. 61-80

E-ISSN 2611-6634 | ISSN 2385-0825
DOI: 10.13133/2611-6634/1415

Research article

Ilaria Stefani[*]

AN OVERVIEW ON DYNAMIC PORTFOLIO ALLOCATION

Abstract

A portfolio allocation problem relies upon the decision process to establish how resources must be allocated among different possible investments. Investors are interested in gaining as much as possible from their investment, but at the same time, they are concerned with the risks they have to face. Investors aim to maximize their returns without exceeding a certain level of risk. Moreover, this behavior has to be mathematically modeled, resorting to the optimal control theory and the maximization of expected utility. This paper reviews the literature on portfolio allocation, to give a complete picture of what has been done, as well as, possible contributions for future research.

Keywords: Asset allocation, Stochastic volatility, Co-jumps, Recursive preferences, Dynamic programming.

---

[*] Department of Methods and Models for Territory, Economics and Finance, Sapienza University of Rome, Italy.

1 Introduction

The first attempt to solve a portfolio allocation problem is due to Markowitz (1952), with the celebrated *mean-variance approach.* The latter continues to be widely applied in several financial frameworks, such as risk management, mainly because of its simplicity, since only the knowledge of the expectation and covariances of random variables is required.

The main drawback of this approach is the static nature of the optimal allocation. Indeed the initial wealth is allocated between different assets at the beginning of the time horizon without allowing for changes in the allocation until maturity. Moreover, this behavior ignores the presence of price volatility, so adopting this type of allocation model could result in large losses.

A solution widely adopted in the literature is to consider continuous-time models for price dynamics: by allowing continuous trading, the investor can immediately react to possible changes in price volatility.

In this view, the pioneering work of Merton (1971) can be considered a starting point for continuous-time portfolio management. In this setting, the asset allocation problem is solved using the stochastic control method, and the optimal portfolio rules are expressed in terms of solutions to the second-order partial differential equation (the so-called *Hamilton-Jacobi-Bellman* (HJB) equation). In this seminal paper, an investor willing to allocate his wealth between stock and a risky asset is taken into account.

The bond price grows at the constant interest rate r, and the stock price dynamic follows the Black-Scholes model

$$\frac{dS_t}{S_t} = \mu dt + \sigma dZ_t,$$

where $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$ and Z is a standard Brownian motion. We assume that the investor's wealth $X = \{X_t\}_{t\in[0,T]}$ is apportioned among the risk-free asset and the stock, so that, for any $t \in [0,T]$, the non-negative process X is governed by the following diffusion dynamics:

$$dX_t = \alpha_t X_t((\mu - r)dt + r)dt + \alpha\sigma X_t dZ_t, \tag{1}$$

where α represents the proportion of wealth invested at time t into the stock and observes that the proportion invested into the riskless asset is equal to $1-\alpha$.

The objective function to maximize over $\alpha \in A$ is of the form

$$U(X_T) = \mathbb{E}\left[\int_0^T f(X_t,\alpha)dt + g(X_T)\right], \tag{2}$$

where f and g are measurable functions on $\mathbb{R}^d \times A$, satisfying suitable integrability conditions and ensuring that the expectation in (2) is well-defined. The dynamic programming method to solve (2) consists in defining the value function, that corresponds to the maximum value for (2) when varying initial states. In symbols, we have

$$v(t,x) = \sup_{\alpha\in A} \mathbb{E}[U(X_T^{t,x})], \quad (t,x) \in [0,T] \times \mathbb{R}^d, \tag{3}$$

where U is an increasing and concave function on $\mathbb{R}^d$.

The corrisponding HJB equation for (3) is

$$\frac{\partial v}{\partial t} + \sup_{\alpha \in \mathbb{R}} \left[(\alpha(\mu - r) + r) x \frac{\partial v}{\partial x}(t,x) + \frac{1}{2} \alpha^2 \sigma^2 x^2 \frac{\partial^2}{\partial x^2}(t,x) \right] = 0, \tag{4}$$

for $(t,x) \in [0,T] \times \mathbb{R}^d$, with terminal condition

$$v(T,x) = U(x), x \in \mathbb{R}^d.$$

Proceeding with the first-order condition for (4), the candidate for the optimal allocation is obtained as follows

$$\hat{\alpha}(t,x) = -\frac{\mu - r}{\sigma^2} \frac{\frac{\partial v}{\partial x}}{x \frac{\partial^2 v}{\partial x^2}}(t,x). \tag{5}$$

Merton has also shown that Equation (4) can be explicitly solved for the special case of the utility function, as when we choose a CRRA (*Constant Relative Risk Aversion*) utility function,

$$\begin{cases} \frac{1}{1-\gamma} x^{1-\gamma}, \; if \; x \geq 0 \\ -\infty, \; if \; x \leq 0 \end{cases}$$

where γ is the coefficient of risk aversion.
It is worth pointing out that the formal derivation of the HJB equation is justified by a verification theorem, which states that, when there exists a smooth solution to the HJB, such a solution coincides with the value function, see Pharm (2007).

An alternative approach to solving optimal allocation problems is the so-called *martingale method* developed by Plisca (1986), Karatzas et al. (1987) and Cox and Huang (1989), such an alternative is based on the transformation of the optimal portfolio allocation problem into a static optimization problem with the determination of the optimal terminal wealth, finding a portfolio strategy that leads to optimal terminal wealth. Several authors extend this approach, we recall that Carr et al. (2001) obtained optimal consumption and investment plans in a complete market when the underlying asset price is a pure jump Lvy process. Schroder and Skiadas (1999) developed the utility gradient approach for computing portfolios and consumption plans that maximize *Stochastic differential utility* function (SDU) and prove the existence, uniqueness, and basic properties of a parametric class of homothetic SDUs. A few years later Schroder and Skiadas (2003) derived closed-form solutions for the optimal consumption and trading strategy in terms of the solution to a single constrained BSDE using the utility gradient approach methodologically. The focus of this paper is the necessary and sufficient first-order conditions of optimality that one would be to solve to compute a solution.

2 Investments under Uncertainty

In a context of uncertainty, an investor chooses among different strategies based on his/her preferences, and this is done in terms of *utility function,* according to the

expected utility criterion, see e.g. Von Neuman and Morgenstern (1947).
If we assume that an investor compares random returns whose probability distributions are known on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we denote by $\succ$ the *preference order* that satisfies the *Von-Neumann Morgenstern criterion,* so that

$$X_1 \succ X_2 \iff \mathbb{E}[U(X_1)] > \mathbb{E}[U(X_2)], \tag{6}$$

where X_1, X_2 are random variables and U is an increasing real-valued function. The choice of the utility function allows defining the concept of *risk aversion* and *risk premium* of the investor. Among the properties that the utility function must have we recall the *concavity* of the utility function. If we have a utility function that satisfies *Jensen's inequality*, then an agent prefers to get the expectation $\mathbb{E}[X]$ of this return (with certainty), that is:

$$U(\mathbb{E}[X]) \geq \mathbb{E}[U(X)], \tag{7}$$

which holds only for concave functions. For a risk-averse agent with concave utility function U we can define the risk premium associated with a random return X he is willing to pay, to get a certain gain:

$$U(\mathbb{E}[X] - \pi) = \mathbb{E}[U(X)]. \tag{8}$$

The quantity $(\mathbb{E}[X] - \pi)$ is called *certainty equivalent* of X, which is smaller than the expectation of X.
It is easy to obtain the *absolute risk aversion* at level x, given by:

$$\alpha(x) = -\frac{U''(x)}{U'(x)}, \tag{9}$$

that is the *Arrow-Pratt* coefficient of absolute risk aversion of U at level x.
In the context of dynamic programming, the choice of the utility function is very relevant. The most preferred class of utility functions is given by HARA (*Hyperbolic Absolute Risk Aversion*) functions, where the inverse of absolute risk aversion is linear in consumption. We mention the quadratic, exponential, and isoelastic preferences. The most popular preference specification is either isoelastic preferences CRRA, or power utility see Hansen and Singleton (1982). The power utility has the following specification:

$$U(x) = \begin{cases} \frac{1}{1-\gamma} x^{1-\gamma}, & if\ x > 0 \\ -\infty, & if\ x \leq 0 \end{cases}, \tag{10}$$

where $\gamma > 0$ and the second part of the utility specification impose a non-negative wealth constraint. The parameter γ plays a crucial role, indeed it is the risk aversion parameter and summarizes the behavior of the consumer toward risk, but at the same time its reciprocal is equal to the *constant elasticity of intertemporal substitution* (EIS) in consumption, and measures how consumption changes in time. So, only one parameter governs both investor risk aversion and EIS, and the latter is independent of the level of consumption.
To give rational support to observable investors' behavior decision-makers might

take into account more complex utility functions, to use a nonlinear aggregator to bring together present and future utility thus dropping the hypothesis of expected utility.

In this perspective, the literature refers to the so-called *stochastic differential utilities* (SDU) introduced by Epstein and Zin (1989), who derive a parametrization of recursive utility in a discrete-time setting, and generalized by Duffie and Epstein (1992) and Fisher and Gilles (1999) in a continuous-time setting.

The Duffie and Epstein (1992) parametrization is

$$J = \mathbb{E}_t \left[\int_t^\infty f(C_s, J_s) ds \right], \tag{11}$$

where $f(C_s, J_s)$ is a normalized aggregator of current consumption and utility with the following form:

$$f(C, J) = \frac{\beta}{1 - \frac{1}{\psi}} (1 - \gamma) J \left[\left(\frac{C}{((1 - \gamma) J)^{\frac{1}{1-\gamma}}} \right)^{1 - \frac{1}{\psi}} - 1 \right], \tag{12}$$

where $\beta > 0$ is the *rate of time preferences*, $\gamma > 0$ is the *coefficient of relative risk-aversion* and $\psi > 0$ is the *elasticity of intertemporal substitution* in consumption. In this setting there is a separation among preferences parameters, so the degree of risk aversion γ is disentangled from the elasticity of intertemporal substitution of consumption EIS. The Epstein-Zin preferences can be seen as a generalization of the standard time additive expected utility function, in that the former collapse to the power utility when the elasticity of intertemporal substitution of consumption EIS equals the reciprocal of the coefficient of relative risk aversion, so when we set $\psi = \frac{1}{\gamma}$.

3 Models for optimal portfolio allocation

3.1 Power Utility

If we focus on portfolio decisions, the literature has produced numerous contributions especially when the investor's decisions are based on his risk aversion, through the isoelastic utility function. Merton (1971) firstly studied optimal portfolio allocation, and derived optimal consumption and investment rules maximizing the expected utility in an economy composed of a riskless asset and a risky stock, when the asset price follows a Geometric Brownian motion. He showed that when investors have some special case of the HARA utility function, the maximization problem can be solved in a closed form.

Subsequently, several authors generalized Merton's work in incomplete markets: we recall, among others, the work of Liu et al. (2003), who used the event-risk framework of Duffie et al. (2000) to provide analytical solutions to optimal portfolio problems by assuming discontinuities in the state variable dynamics. In particular, Liu et al. (2003) provided two examples to illustrate their results. In the first one, they consider a model where the risky asset follows a jump-diffusion process with deterministic jump size, assuming constant volatility. They find that an investor would hold less of the risky asset when the price jumps occur. In the second example the authors consider a model where both the risky asset $S = \{S_t\}_{t\in[0,T]}$ and its return volatility $V = \{V_t\}_{t\in[0,T]}$ follow jump-diffusion processes with deterministic jump sizes, in particular, consider the following dynamics:

$$\begin{cases} dS_t = (r + \eta V_t - J\lambda V_t)S_t dt + \sqrt{V_t} S_t dZ_t^{(1)} + JS_{t^-} dN_t \\ dV_t = (k\theta - kV_t - \xi\lambda V_t)dt + \sigma\sqrt{V_t)} dZ_t^{(2)} + \xi dN_t \end{cases}, t \in [0,T] \tag{13}$$

where $Z_t^{(1)}$and $Z_t^{(2)}$ are standard Brownian motions with correlation ρ. V_t is the instantaneous variance of diffusive returns and N_t is a Poisson process with stochastic arrival intensity λV_t, assuming λ to be nonnegative. The parameter $\theta > 0$ is the long-run mean, $k > 0$ is the mean-reversion rate, $\sigma > 0$ is the vol-of-vol coefficient and η is the market risk premium. Finally, J, $\xi > 0$ are the jump sizes, assumed to be constant.
Defining the indirect utility function W as

$$W(X, V, t) = \max_{\phi_s, t\le s\le T} \mathbb{E}_t \left[U(X_T)\right], \tag{14}$$

we have the following HJB equation for the indirect utility function W:

$$\max_{\phi}\Bigg(\frac{\phi^2 X^2 V}{2} W_{XX} + \phi\rho\sigma XVW_{XV} + \frac{\sigma^2 V}{2} W_{VV} + (r + \phi(\eta - J\lambda)V)XW_X + \tag{15}$$
$$(k\theta - kV - \xi\lambda V)W_V + \lambda V(\mathbb{E}[W(X(1+\phi J), V+\xi, t)] - W) + W_t\Bigg) = 0$$

where W_X, W_V and W_t denote the derivatives of $W(X, V, t)$ with respect to X, V and t and similarly for the higher derivatives. Proceeding with a *Guess and Verify* approach, the problem is solved assuming that the indirect utility function is of the form

$$W(X, V, t) = \frac{1}{1-\gamma} X^{1-\gamma} \exp\{F_t V_t + G_t\}, \tag{16}$$

where F_t and G_t are deterministic functions of time .
Depending on the model F_t and G_t can be solved in closed form or numerically, the latter case is that of this model. This model led to the following expression for optimal portfolio weights:

$$\phi^* = \frac{\eta - J\lambda}{\gamma} + \frac{\rho\sigma F_t}{\gamma} + \frac{\lambda J}{\gamma}(1 + J\phi^*)^{-\gamma} e^{\xi F_t}, \tag{17}$$

where F is the solution to the following ordinary differential equation

$$\dot{F}_t + \frac{\sigma^2 F_t^2}{2} + (\phi^* \rho\sigma(1-\gamma) - k - \xi\lambda)F_t + \tag{18}$$
$$\left(\frac{\gamma(\gamma-1)\phi^{*2}}{2} + (\eta - J\lambda)(1-\gamma)\phi^* + \lambda M_2 - \lambda\right) = 0,$$

with $\dot{F}_t = \frac{\partial F}{\partial t}$, Liu et al. (2003) for further technical details.
It is possible to note that the investment horizon affects optimal portfolio weights through the hedging component of the demand for the risky asset ($\frac{\rho\sigma F_t}{\gamma}$) and from the static hedging component ($e^{\xi F_t}$) in Equation (18). Moreover, the authors found that volatility jumps can have a significant effect on optimal portfolio: in presence of jumps in volatility, the investor increases the optimal allocation in the risky asset. This means that volatility jumps have a compensating effect concerning price jumps.

Liu (2007) solved dynamic portfolio choice problems using affine models to face stochastic volatility. Three applications are presented: the first one is the bond portfolio selection problem when bond returns are described by quadratic term structure models; the second one is the stock portfolio selection problem when volatility is stochastic as in the Heston model; the last application is a portfolio selection problem in the incomplete market when the interest rate is stochastic and stock returns have stochastic volatility.
Very relevant to understand the subsequent works is to focus on the second application, the following dynamics are considered:

$$\left\{ \begin{array}{l} dS_t = (r + \eta V_t)S_t dt + \sqrt{V_t} S_t dZ_t^{(1)} \\ dV_t = (k\theta - kV_t)dt + \sigma\sqrt{V_t} dBZ_t^{(2)} \end{array} \right. , t \in [0, T] \tag{19}$$

where $S = \{S_t\}_{t\in[0,T]}$ is the stock price, $V = \{V_t\}_{t\in[0,T]}$ is the volatility and $\eta V_t = \mu - r$ is the risk premium.
Associated with the system (19) we have the following HJB equation:

$$\max_{\phi}\left(\frac{\phi^2 X^2 V}{2} W_{XX} + \phi\rho\sigma XVW_{XV} + \frac{\sigma^2 V}{2} W_{VV} + (r + \phi\eta VXW_X + \right. \tag{20}$$
$$\left. (k\theta - kV)W_V + W_t \right) = 0,$$

where $W_X = \frac{\partial W}{\partial X}$, $W_V = \frac{\partial W}{\partial V}$, $W_t = \frac{\partial W}{\partial t}$, $W_{XX} = \frac{\partial^2 W}{\partial X^2}$, $W_{VV} = \frac{\partial^2 W}{\partial V^2}$ and $W_{XV} = \frac{\partial W}{\partial X \partial V}$.
The problem is solved assuming Equation (16) as an indirect utility function.
The optimal stock portfolio weight ϕ_s^* is given by

$$\phi_s^* = \frac{1}{\gamma}\eta + \rho\sigma F_t, \tag{21}$$

where

$$F_t = \left\{ \begin{array}{l} -\frac{2[exp(k_2 t) - 1]}{(k_1 + k_2)[exp(k_2 t) - 1] + 2k_2}\delta_v \quad if \quad k_2^2 \geq 0 \\ -\frac{2}{k_1 + \zeta\frac{cos(\zeta t/2)}{sin(\zeta t/2)}}\delta \quad if \quad \zeta^2 \geq 0 \end{array} \right. , \tag{22}$$

and

$$k_1 = k - \frac{1-\gamma}{\gamma}\alpha\sigma\rho, k_2 = \sqrt{k_1^2 + 2\delta[\rho^2 + \gamma(1-\rho^2)]\sigma^2}\delta = -\frac{1-\gamma}{2\gamma^2}\alpha^2, \zeta = -ik_2.$$

The stock portfolio weight in the stochastic volatility model is a nonmonotonic function of risk aversion. Indeed the myopic component decreases when risk aversion increases. A surprising and criticized feature of the optimal portfolio weight is that it is independent of the variance V. One might be expected the agent to hold more stocks when the volatility is low and less when the volatility is high, but this is true if the risk premium is independent of the volatility: in this model, the risk premium is proportional to the conditional variance. Hence, when the variance is high, the risk premium is also high, and vice versa.

We observe that the aforementioned findings are not related to a specific verification theorem, so we do not know under which conditions the candidate for the optimal portfolio strategy is the unique optimal solution to the allocation problem.

The work by Kraft (2005) filled this gap, by proving a verification result within the stochastic volatility framework, the author also derived the parameters' conditions ensuring well-defined candidates for the solution of the problem, a condition that ensures that these are indeed the solutions of the optimal portfolio process given the value function.

In contrast, Pharm (2022) and Fleming and Hernandez-Hernandez (2003) derived explicit verifications results proving that their portfolio strategies are indeed optimal. In particular, Pharm (2022) considered a multidimensional model for securities with stochastic volatilities, assuming certain Lipschitz conditions for coefficients. Such a choice excluded *de facto* the Heston model.

On the other hand, Fleming and Hernandez-Hernandez (2003) assumed the asset volatility to be a function σ depending on a state process with constant volatility, with boundedness assumptions for σ and σ'. Also, in this case, the Heston model is not included.

Buraschi et al. (2010) developed a new framework for multivariate intertemporal portfolio choice, in which the correlations across asset classes and volatilities are stochastic. In this setting, volatilities and correlations are conditionally correlated with returns. To model stochastic variance-covariance risk, the authors specified the covariance matrix process as a Wishart diffusion process. More precisely, we refer to the bi-dimensional case, so that the dynamics of the price vector $S_t = (S_1, S_2)^T$ are described by the bivariate stochastic differential equation:

$$\begin{cases} dS_t = diag(S_t)\left[(r + \mathbf{\Lambda}(\Sigma, t))dt + \Sigma_t^{1/2} dZ_t^{(1)}\right] \\ d\Sigma_t = \left[\Omega\Omega' + K\Sigma_t + \Sigma_t K'\right]dt + \Sigma_t^{1/2} dZ(t)^{(2)} Q + Q'(dZ_t^2)'\Sigma_t^{1/2} \end{cases}, t \in [0, T] \quad (23)$$

where $diag(S_t)$ is the square matrix with S_t in the diagonal and 0 on the off-diagonal elements, $\mathbf{\Lambda}(\Sigma, t))$ is a vector of possibly state-dependent risk premia. Here the latter is assumed to be a constant market price of variance-covariance risk, namely $\mathbf{\Lambda}(\Sigma, t)) = \Sigma(t)\eta$ with $\eta = (\eta_1, \eta_2)^T \in \mathbb{R}^2$. Finally, the processes $Z_t^{(1)} \in \mathbb{R}^{2\times 1}$, and $Z_t^{(2)} \in \mathbb{R}^{2\times 2}$ are matrix Wiener processes, while K, Q, Ω are 2×2 square matrix (with Ω invertible). One property that the model must satisfy is $\Omega\Omega' = \varphi QQ'$ with $\varphi \in \mathbb{R}$ and $\varphi > N - 1$.

The Wiener processes determining shocks in prices S_t and in the variance-covariance matrix Σ_t are correlated according to

$$Z_t^{(1)} = Z_t^{(2)}\rho + \sqrt{1-\rho'\rho}Z_t^{(3)}, \tag{24}$$

where $\rho = (\rho_1, \rho_2)$ is a vector of correlation parameters, and $Z_t^{(3)}$ is a two-dimensional standard Brownian motion, independent of $Z_t^{(2)}$.
The optimal weights obtained in such a framework are:

$$\phi^* = \frac{\alpha}{\gamma} + 2F_t Q'\rho, \tag{25}$$

with F_t being the solution to the following equation

$$\dot{F}_t + \Gamma' F_t + F_t\Gamma + 2F_t\Lambda F_t + C = 0, F_T = 0. \tag{26}$$

The optimal weight in (25) is the sum of a *myopic demand* and *hedging demand* of the risky asset, as already highlighted in Liu (2007) in the univariate case, and it is independent of the variance. Moreover, the authors pointed out that the hedging demand is typically larger than in univariate models, and it includes a significant covariance hedging component.

In a similar context, Jin and Zhang (2012) examined a multi-asset model with constant volatility when the price can jump. Differently from other authors, such as Das and Uppal (2004) and Ait-Sahalia et al. (2009), who focused on one type of jump and assumed constant coefficients in their dynamics, Jin and Zhang (2012) includes multiple types of jumps and a large number of assets and state variables. Furthermore, this paper also incorporates model uncertainty into the portfolio choice problem, tackling the problem by focusing on ambiguity aversion to inaccurate estimates of parameters associated with jumps. The theoretical results showed that ambiguity aversion increases the effect of jumps for a risk-averse investor, thus the exposure to jumps is smaller in the case of an investor not neutral to ambiguity.

The evidence highlighted in the works mentioned above shows once again that the optimal weights result to be independent of the instantaneous volatility when the dynamics of the risky asset are affine, implying a static allocation mechanism.

To overcome this drawback, Oliva and Ren (2018) study a continuous time optimal portfolio allocation with volatility and co-jump risk in a multi-asset framework. They deviate from affine models by specifying a Wishart jump-diffusion for the co-precision (the inverse of the covariance matrix), generalizing the dynamics in Chacko and Viceira (2005) to a multivariate setting. They solved the optimal portfolio problem, providing an exact solution to this problem in the absence of jumps and an approximated solution in the presence of co-jumps. More precisely, it is assumed there exist N risky assets, whose prices are given by $S_t = (S_{t,1}, \cdots, S_{t,N})^T \in \mathbb{R}^{\mathbb{N}\times 1}$ with dynamics

$$\left\{ \begin{array}{l} dS_t = diag(S_t)\left[(\eta dt + \sqrt{Y_t^{-1}}dZ_t^{(1)} + JdN(\lambda)_t\right] \\ dY_t = \left[\Omega\Omega' + KY_t + Y_tK'\right]dt + \sqrt{Y_t}dZ_t^{(2)}Q + Q'(dZ_t^2)'\sqrt{Y_t} + \xi(Y_t)dN(\lambda)_t \end{array} \right., t \in [0,T] \tag{27}$$

where $diag(S_t)$ is the diagonal matrix with S_t in the diagonal and 0 on the off-diagonal elements, $\eta, J \in \mathbb{R}^{N\times 1}$, $Z_t^{(1)} \in \mathbb{R}^{N\times 1}$ and $Z_t^{(2)} \in \mathbb{R}^{N\times N}$ are matrix Wiener processes, $K, Q, \xi \in \mathbb{R}^{N\times N}$ while $N(\lambda)_t$ is a non-compensated Poisson process with intensity $\lambda \in \mathbb{R}$. Furthermore Ω is a symmetric, positive definite and invertible matrix such that $\Omega\Omega' = \varphi QQ'$ with $\varphi \in \mathbb{R}$ and $\varphi > N-1$. The Wiener processes determining shocks in prices and variance-covariance matrix $\Sigma_t = Y_t^{-1}$ are correlated according to

$$Z_t^{(1)} = Z_t^{(2)}\rho + \sqrt{1-\rho'\rho}Z_t^{(3)}, \tag{28}$$

where $\rho \in \mathbb{R}^{N\times 1}$, $Z_t^{(3)} \in \mathbb{R}^{N\times 1}$is a Wiener process and the elements of $Z_t^{(3)}$ and $Z_t^{(2)}$ are all independent among them.

The approximated HJB for the investment problem is:

$$\max_{\phi}\Bigg(W_t + [\phi_t'(\eta - r\mathbf{1}) + r]X_tW_X + Tr([\Omega\Omega' + KY_t + Y_tK']\nabla W) + \frac{1}{2}X_t^2\phi'Y_t^{-1}\phi_t \\ W_{XX} + (2\phi_t'\nabla Q'\rho W_X)\,X + \frac{1}{2}Tr\,(4Y_t\nabla Q'Q\nabla)\,W + \lambda\mathbb{E}[W(X_t + \phi'JX_t, Y_t + \\ \xi) - W(X_t, Y_t)]\Bigg), \tag{29}$$

where $\nabla := \left(\frac{\partial}{\partial Y_{i,j}}\right)_{1\leq i,j\leq N}$.

The value function associated with Equation (29) is given by:

$$W(t, Y_t, X_t) = \exp\{Tr(F_tY_t) + G_t\}\frac{X_t^{1-\gamma}}{1-\gamma}, \tag{30}$$

where Tr represents the trace of a square matrix. It's important to note that to recover a viable HJB equation in this realistic case the authors follow an approximated approach. The first approximation consists in assuming that the jump matrix in the co-precision ξ is constant, not depending on Y_t. The second approximation consists of the linearization of the jump term appearing in the HJB equation, using second-order Taylor expansion:

$$(1 + \phi_t'J)^{1-\gamma} = 1 + (1-\gamma)\phi_t'J + o((\phi_t'J)^2), \tag{31}$$

and the approximation consists in not consider the $o((\phi_t'J)^2)$ term in the HJB equation.

The solution to the optimal allocation is dynamic consistently with the well-known Markowitz economic intuition, in that the optimal weight is proportional to the instantaneous co-precision:

$$\phi^* = Y_t\left[\frac{(\eta - r1) + 2F_tQ'\rho + \lambda\mathbb{E}\left[e^{Tr(F_t\xi)}J\right]}{\gamma}\right] =: Y_tB_t, \tag{32}$$

where F_t solves

$$\dot{F}_t + (1-\gamma)(\eta) - r1B_t' + F_tK + K'F_t - \frac{\gamma(1-\gamma)}{2}B_tB_t' + 2(1-\gamma)F_tQ'\rho B_t' \\ + 2F_tQ'QF_t + \lambda(1-\gamma)\mathbb{E}\left[e^{Tr(F_t\xi)}J\right]B_t' = 0, F_T = 0. \tag{33}$$

The optimal portfolio weights consist of three terms: the first one is the *myopic demand* component, which is now proportional to inverse volatility and so takes the typical form of standard mean-variance allocation; the second one is the *intertemporal hedging demand* term, which depends on the correlation coefficients and co-precision. Finally, the third term assumes the meaning of an illiquidity term.

Differently, Zhou et al. (2019) solves the dynamic portfolio allocation considering an AR(1)-GARCH(1,1)-ARJI model to describe the asset returns which enables capturing the dynamics process of both volatility and jump intensity. The authors find that both initial jump intensity and jump persistence are important for the investor's optimal portfolio decision.

More recently Jin et al. (2021) formulates a portfolio choice problem in a multi-asset market characterized by ambiguous jumps and constant volatility. In the paper, therefore, a portfolio choice problem with ambiguity and ambiguity aversion in a continuous-time incomplete financial market is considered. Recall that ambiguity is uncertainty that cannot be measured by a single probability measure. The authors investigate the impact of tail risk on portfolio selection. The optimal portfolio is solved in closed form through a decomposition approach, with two different jump size distributions. They also show that underestimating tail risk might result in a sizeable wealth loss in the presence of jump ambiguity. Furthermore, they confirm that ambiguity-averse investors reduce more of their jump exposure if the jump distribution exhibits a fatter left tail.

So far we have assumed an incomplete market setting. To complete the market, we can add some financial securities susceptible to the whole range of risk components. The natural choice is to include derivative contracts on the underlying portfolio equity. Among the numerous references we first focus on Liu and Pan (2003) who studied the impact of options on wealth allocation in a stochastic volatility framework considering only jumps in price. In particular, they consider the following dynamic for the risky asset $S = \{S_t\}_{t\in[0,T]}$:

$$\begin{cases} dS_t = (r + \eta V_t + J(\lambda - \lambda^{\mathbb{Q}})V_t)S_t dt + \sqrt{V_t} S_t dZ_t^{(1)} + J S_{t-}(dN_t - \lambda V_t dt) \\ dV_t = k(\theta - V_t)dt + \sigma\sqrt{V_t}\left(\rho dZ_t^{(1)} + \sqrt{1-\rho^2} dZ_t^{(2)}\right) \end{cases}, t \in [0, T] \tag{34}$$

where $Z_t^{(1)}$ and $Z_t^{(2)}$ are standard Brownian motions, and N is a pure-jump process. All the random shocks are assumed to be independent. The instantaneous variance process $V = \{V_t\}_{t\in[0,T]}$ is a stochastic process with long-run mean $\theta > 0$, mean-reversion rate $k > 0$, and vol-of-vol coefficient $\sigma \geq 0$. In this setting, the agent is interested in investing not only in stock and riskless bonds but also in derivatives. More specifically the derivatives involved are those providing different exposures to the three fundamental risk factors in the economy. The market can be completed once we introduce enough *non-redundant* derivatives $O_t^i = g^i(S_t, V_t)$ for $i = 1 \cdots N$

and the price dynamics for the i-th derivative security is:

$$
\begin{aligned}
dO_t^{(i)} =& rO_t^i dt + (g_s^{(i)} S_t + \sigma\rho g_v^{(i)})(\eta V_t dt + \sqrt{V_t} dB_t) + \sigma\sqrt{1-\rho^2} g_v^{(i)} (\epsilon V_t dt + \sqrt{V_t} dZ_t^{(1)} \\
& + \Delta g^{(i)} (\lambda - \lambda^{\mathbb{Q}}) V_t dt + dN_t - \lambda V_t dt,
\end{aligned} \tag{35}
$$

where $g_s^{(i)}$ and $g_v^{(i)}$ measure the sensitivity of the $i-$th price to infinitesimal changes in the stock price and volatility, respectively, and where $\Delta_g^{(i)}$ measures the change in the derivative price for each jump in the underlying stock price. To solve the investment problem, the authors define the indirect utility function

$$
W(t, X_t, V_t) = \frac{X_t^{1-\gamma}}{1-\gamma} \exp\{\gamma F(T-t) V_t + \gamma G(T-t)\}. \tag{36}
$$

which satisfies the following HJB equation,

$$
\begin{aligned}
\max_{\phi_t, \nu_t} \Bigg(& W_t + X_t W_X \left(r_t + \theta_t^B \eta V_t + \theta_t^Z \epsilon V_t - \theta_t^N J \lambda^Q V_t\right) + \frac{1}{2} X_t^2 W_{XX} V_t \left((\theta_t^B)^2 + (\theta_t^Z)^2\right) \\
& + \lambda V_t \Delta W + k(\theta - V_t) W_V + \frac{1}{2} \sigma V_t W_{VV} + \sigma V_t X_t W_{WV} \left(\rho \theta_t^B + \sqrt{1-\rho.^2} \theta_t^Z\right) \Bigg) \\
& = 0,
\end{aligned} \tag{37}
$$

where $\Delta W = W(t, X_t(1 + \theta^N J), V_t) - W(t, X_t, V_t)$ denotes the jump in the indirect utility function W for given jumps in the stock price, and where W_t, W_X and W_V denotes the derivatives of $W(t, X, V)$ with respect to t, X and V, similar notations for higher derivatives. The authors solve the optimization problem in closed form and the optimal portfolio weights on the risk factors B (bond), Z(stock), and N (derivatives) are given by:

$$
\begin{aligned}
\theta_t^{*B} &= \frac{\eta}{\gamma} + \sigma\rho F(\tau), \\
\theta^{*Z} &= \frac{\epsilon}{\gamma} + \sigma\sqrt{1-\rho^2} F(\tau,) \\
\theta^{*N} &= \frac{1}{J}\left(\left(\frac{\lambda}{\lambda^{\mathbb{Q}}}\right)^{1/\gamma} - 1\right),
\end{aligned} \tag{38}
$$

where η and ϵ are the risk premia and

$$
F(\tau) = \frac{exp(k_2\tau) - 1}{2k_2 + (k_1 + k_2)(exp(k_2\tau) - 1)} \delta, \tag{39}
$$

with

$$
\begin{aligned}
\delta &= \frac{1-\gamma}{\gamma}(\eta^2 + \epsilon^2) + 2\lambda^{\mathbb{Q}} \left[\left(\frac{\lambda}{\lambda^{\mathbb{Q}}}\right) + \frac{1}{\gamma}\left(1 - \frac{\lambda}{\lambda^{\mathbb{Q}}}\right) - 1\right], \\
k_1 &= k - \frac{1-\gamma}{\gamma}\left(\eta\rho + \epsilon\sqrt{1-\rho^2}\right)\sigma, \\
k_2 &= \sqrt{k_1^2 - \delta\sigma^2}.
\end{aligned}
$$

The optimal exposure to the three risk factors, according to Liu et al. (2003) for the incomplete market setting, does not depend on instantaneous volatility. The authors showed as derivatives extend the risk and return tradeoffs associated with stochastic volatility and price jumps. In particular, they illustrated two significant examples. In the first example, they focused on the role of derivatives as a vehicle for volatility risk and as result, the optimal portfolio weight in derivative security depends explicitly on the sensitivity of the derivative to volatility. The second example is the role of derivatives as a vehicle to disentangle jump risk from diffusive risk.

Later Branger et al. (2007) extended the Liu and Pan (2003) framework by also considering discontinuities in the stochastic process that governs volatility. They showed that the demand for jump risk includes a hedging component which is not present in the models without volatility jump, this is the main difference with the previous setting. More precisely, we have

$$\theta^{*N} = \frac{1}{J}\left[\left(\frac{\lambda}{\lambda^{\mathbb{Q}}}\right)^{1/\gamma} - 1\right] + \frac{1}{J}\left(\frac{\lambda}{\lambda^{\mathbb{Q}}}\right)\left[e^{F(\tau)\xi} - 1\right], \tag{40}$$

with ξ volatility jump size assumed to be constant, and consequently this impacts also on $F(\tau)$, equation (39), while the other exposures $(\theta^{*B}, \theta^{*Z})$ are the same as Equation (38) of Liu and Pan (2003).
Moreover, the authors showed how the volatility jump magnitude has a significant impact on the optimal portfolio. They analyzed the distribution of terminal wealth for an investor who uses the wrong model by ignoring volatility jumps or including wrong estimates of such jumps.

In this context, we also mention Escobar et al (2015), who determine the optimal portfolio for an ambiguity-averse investor when stock price follows a stochastic volatility jump-diffusion process (considering only a jump in the stock dynamics), and when the investor can have different levels of uncertainty about diffusion parts of the stock and its volatility. The authors illustrate that the optimal exposures to stock and volatility risks are significantly affected by ambiguity aversion to the corresponding risk factor only. Moreover, they also show that volatility ambiguity has a smaller impact in incomplete markets. As an extension, Cheng and Escobar-Anel (2021) consider an optimal allocation problem with both risk and ambiguity aversion under a 4/2 model. The numerical analysis finds that *wealth-equivalent losses* (WELs) from ignoring uncertainty or market completeness are moderate, while the WELs for investors who follow different models such as Heston or Merton (geometric Brownian motion) is quite substantial.

Unlike previous works in Ilhan and Sicar (2005) the investors maximize *expected exponential utility* function of terminal wealth and restrict a static position in derivative securities. The main result is that in a general incomplete arbitrage-free market there exists a unique optimal strategy for the investor.

Haugh and Lo (2001) construct a buy-hold portfolio of stocks, bonds, and options that involves no trading once set at the beginning of the investment horizon, and solve this problem for several combinations of preferences as CRRA and CARA (*Constant Absolute Risk Aversion*) preferences and different return dynamics (*Geometric Brownian motion*, the *Ornstein Uhlenbeck process* and a *Bivariate Linear Diffusion process*). The authors show that under certain conditions a portfolio consisting of just a few options is an excellent substitute for more complex dynamic policies.

3.2 Non-expected utility: EpsteinZin preferences

As an alternative to power utilities, a strand of extant literature has focused on *Epstein-Zin preferences*, since they include the effect of both risk aversion, and separate EIS from the coefficient of relative risk aversion. The power utility functions restrict risk aversion to be the reciprocal of the elasticity of intertemporal substitution and this does not reflect the empirical evidence that has shown how these parameters have very different effects on optimal consumption and portfolio choice, as highlighted in Chacko and Viceira (2005). The latter examines the optimal consumption and portfolio-choice problem of long-horizon investors in an incomplete market setting, by first introducing *precision* process, intended as the inverse of volatility, to obtain dynamic optimal portfolio rules. The authors consider recursive utility over consumption and derive an analytic expression for the optimal consumption and portfolio policies. The latter are exact when an investor has unit elasticity of intertemporal substitution of consumption, and approximate otherwise. To simplify the analysis, the market is assumed to be made of only two assets, the first one is the riskless asset with dynamics:

$$\frac{dB_t}{B_t} = rdt, \tag{41}$$

where r is the risk-free rate. The second one is stock, with the following dynamics :

$$\begin{cases} dS_t = \mu S_t dt + \sqrt{y_t^{-1}} S_t dZ_t^{(1)} \\ dy_t = k(\theta - y_t)dt + \sigma\sqrt{y_t} dZ_t^{(2)} \end{cases}, t \in [0, T] \tag{42}$$

where y_t is the instantaneous precision of the risky asset return that follows a mean-reverting process, with $\frac{1}{v_t} = y_t$, being v_t the variance process, and $k, \theta, \sigma > 0$. Moreover in such a setting shocks to precision are correlated with the instantaneous returns on the risky asset, with $\langle dZ_t^{(1)} dZ_t^{(2)} \rangle = \rho dt$.
The HJB equation for this problem is:

$$\sup_{\phi_t, C} \Bigg(f(C_s, J_s) + [\phi_t(\mu - r)X_t + rX_t - C_t]\, W_X + \frac{1}{2}\phi_t^2 X_t^2 W_{XX} \frac{1}{y_t} + k(\theta - y_t)W_y + \\ \frac{1}{2}\sigma^2 W_{yy} y_t + \rho\sigma\phi_t X_t W_{Xy} \Bigg) = 0, \tag{43}$$

where $f(C, J) = \beta(1-\gamma)W\left[\log(C) - \frac{1}{1-\gamma}\log((1-\gamma)W)\right]$ is the aggregator for $\psi = 1$ and subscripts on W denote partial derivatives. When $\psi = 1$, there is an exact analytical solution to the optimization problem with value function given by:

$$W(X_t, y_t) = \frac{X_t^{1-\gamma}}{1-\gamma}\exp\{Fy_t + G\}, \tag{44}$$

where F_t and G_t are deterministic functions to be determined. This value function implies the following optimal consumption and portfolio rules:

$$\begin{cases} \frac{C_t}{X_t} = \beta \\ \phi_t^* = \frac{1}{\gamma}(\mu - r)y_t + \left(1 - \frac{1}{\gamma}\right)(-\rho)\sigma F y_t \end{cases}, \tag{45}$$

with $\beta > 0$ rate of time preferences, $\gamma > 0$ is the coefficient of relative risk aversion, and F_t, G_t are given by the solution of the following system:

$$aF^2 + bF + c = 0, \tag{46}$$

$$(1-\gamma)(\beta\log\beta + r - \beta) - \beta G + k\theta F = 0 \tag{47}$$

where

$$a = \frac{\sigma^2}{2\gamma(1-\gamma)}[\gamma(1-\rho^2) + \rho^2], b = \frac{\rho\sigma(\mu - r)}{\gamma}, c = \frac{(\mu - r)^2}{2\gamma}. \tag{48}$$

When $\psi \neq 1$ the HJB is still given by (43) but the first-order condition for consumption is different due to the different form of the aggregator, as in equation (12). So, in this case, guessing that

$$W(X_t, y_t) = I(y_t)\frac{X_t^{1-\gamma}}{1-\gamma} \tag{49}$$

with the transformation $I = H^{\frac{-1-\gamma}{1-\psi}}$ and replace into HJB equation, also replacing the expression for ϕ_t, the authors obtain a *non-homogeneous* ordinary differential equation:

$$\begin{aligned} &-\beta^{\psi}H^{-1} + \phi\beta + \frac{(1-\psi)(\mu - r)^2}{2\gamma}y_t - \frac{\rho\sigma(\mu - r)(1-\gamma)}{\gamma}\frac{H_y}{H}y_t + r(1-\psi) + \\ &\frac{\rho^2\sigma^2(1-\gamma)^2}{2\gamma(1-\psi)}\left(\frac{H_y}{H}\right)^2 y_t - \frac{H_y}{H}k(\theta - y_t) + \frac{\sigma^2}{2}\left(\frac{1-\gamma}{1-\psi} + 1\right)\left(\frac{H_y}{H}\right)^2 y_t\frac{\sigma^2}{2}\frac{H_{yy}}{H}y_t = 0, \end{aligned} \tag{50}$$

the solution is obtained by approximating the term

$$\beta^{\psi}H^{-1} = \exp\{c_t - x_t\}, \tag{51}$$

where $c_t - x_t = \log(C_t/X_t)$ and using a first-order Taylor expansion of $\exp\{c_t - x_t\}$:

$$\beta^{\psi}H^{-1} \approx h_0 + h_1(c_t - x_t),$$

where $h_1 = \exp\overline{c - x}$ and $h_0 = h_1(1 - \log h_1)$. Substituting Equation (51) in the first term of Equation (50), the resulting ODE has a solution of the form $H = \exp\{F_1 y_t + G_1\}$. This solution implies the following value function

$$W(X_t, y_t) = \frac{X_t^{1-\gamma}}{1-\gamma}\exp\left\{-\left(\frac{1-\gamma}{1-\psi}\right)(F_1 y_t + G_1)\right\}. \tag{52}$$

In this latter case the optimal consumption is equal to $\frac{C_t}{X_t} = \beta^{\psi} e^{-F_1 y_t - G_1}$, F_1 and G_1 are given by solution to a system similar to (46), see Appendix A. of Chacko and Viceira (2005) for more details.
The optimal portfolio rule has two components, namely a *myopic portfolio demand*, and a Merton's *intertemporal hedging demand*. Both components are linear functions of precision. The optimal consumption wealth ratio is invariant to changes in volatility if $\psi = 1$, while it is an affine function of instantaneous precision when $\psi \neq 1$. In a similar framework, Faria and Correia-da-Silva (2016) extended the model of Chacko and Viceira (2005) for optimal dynamic portfolio choice, introducing ambiguity in stochastic investment opportunity set, showing a small impact of the elasticity of intertemporal substitution of consumption (EIS) on optimal allocation. Since standard verification results are not applicable (Duffie and Epstein (1992)), due to the non-Lipschitzianity of Epstein-Zin preferences, Kraft et al. (2013) provided a suitable verification theorem for the associated HJB equation. This paper contributes to providing new explicit solutions to the HJB equation with recursive utility for a non-unit EIS. Those results represent the first explicit benchmark for the *Cambell-Shiller approximation*, used by Chacko and Viceira (2005) in their approximation. Kraft et al. (2016) extended their previous work and provided the existence and uniqueness of solutions of HJB equation by exploiting fixed point arguments, and developed a fast and accurate numerical method for computing both indirect utility and optimal strategies.

Xing (2017) studied an investment problem via backward stochastic differential equations considering a multi-asset model in which the assets follow a Geometric Brownian motion and volatility is constant, focusing on the empirically relevant specification where both risk aversion and EIS are greater than one. The utility specification makes the optimization problem very difficult to solve since the Epstein-Zin aggregator is not Lipschitzian.

In a complete market setting, we refer to Hsuku (2007), where a recursive utility function defined on intermediate consumption (rather than terminal wealth) is maximized, to reflect the realistic behavior of an investor who saves money for the future. This setting is based on a general assumption according to which expected stock returns are affine functions of volatility. The economy is formed by a riskless bond

$$\frac{dB_t}{B_t} = r dt, \tag{53}$$

where r is the risk-free rate and a risky stock with the following dynamic:

$$\begin{cases} \frac{dS_t}{S_t} = \mu dt + \sqrt{V_t} dZ_t^{(1)} \\ dV_t = k(\theta - V_t)dt + \sigma\sqrt{V_t}\rho dZ_t^{(1)} + \sqrt{1-\rho^2} dZ_t^{(2)} \end{cases}, t \in [0, T] \tag{54}$$

where $Z_t^{(1)}$ and $Z_t^{(2)}$ are standard Brownian motions. In such a market, the authors refers to Liu and Pan (2003), so that the non-redundant derivative $O_t = g(S_t, V_t)$

at the time $t \in [0,T]$ has the following dynamics:

$$dO_t = \Big[(\mu - r)(g_s S_t + \rho\sigma g_v) + \eta\sigma\sqrt{1-\rho^2} g_v + rO_t\Big]\, dt + (g_s S_t + \rho\sigma g_v)\sqrt{V_t} Z_t^{(1)} +$$
$$(\sigma\sqrt{1-\rho^2} g_v)\sqrt{V_t} dZ_t^{(2)}, t \in [0,T] \tag{55}$$

where η determines the stochastic volatility risk premium, g_s and g_v are measures of derivative price sensitivities to small changes in the underlying stock price and volatility, respectively. When $\psi \to 1$ the optimal portfolio weights in stock and derivatives, and the optimal consumption wealth ratio is:

$$\frac{C_t}{X_t} = \beta, \tag{56}$$
$$\phi_t^* = \frac{1}{\gamma}\frac{\mu - r}{V_t} + \frac{1}{\gamma}\left(Q_1 + Q_2\right)\frac{1}{V_t})\rho\sigma - \frac{1}{\gamma}\frac{\eta}{V_t}\frac{(g_s S_t + \rho\sigma g_v)}{\sigma\sqrt{1-\rho^2} g_v} - \frac{1}{\gamma}\left(Q_1 + Q_2\frac{1}{V_t}\right) \tag{57}$$
$$\frac{(g_s S_t + \rho\sigma g_v)}{g_v},$$
$$\nu^* = \frac{1}{\gamma}\frac{\lambda}{V_t}\frac{1}{\sigma\sqrt{1-\rho^2}(g_v/O_t)} + \frac{1}{\gamma}\left(Q_1 + Q_2\frac{1}{V_t}\right) O_t/g_v, \tag{58}$$

with Q_1 and Q_2 that solve the following equations:

$$\left(\frac{1}{2}\frac{1}{1-\gamma}\sigma^2 + \frac{1}{2}\frac{1}{\gamma}\sigma^2\right) Q_2^2 + \left[\frac{1}{\gamma}\sigma(\mu - r)\rho\lambda\sqrt{1-\rho^2} + \frac{1}{1-\gamma}k\theta - \frac{1}{2}\frac{1}{1-\gamma}\sigma^2\right] Q_2 + \tag{59}$$
$$\frac{1}{2}\frac{1}{\gamma}[(\mu - r)^2 + \lambda^2] = 0, \left(\frac{1}{2}\frac{1}{1-\gamma}\sigma^2 + \frac{1}{2}\frac{1}{\gamma}\sigma^2\right) Q_1^2 - \left(\frac{1}{1-\gamma}\beta + \frac{1}{1-\gamma}k\right) Q_1 -$$
$$\frac{1}{1-\gamma}\beta\frac{1}{\theta}Q_2 = 0.$$

For the general case $\psi \neq 1$ there is no exact solution. To find an approximate solution the authors apply the approximation proposed in Chacko and Viceira (2005). Equation (56) demonstrates the invariance of the optimal log consumptionwealth ratio to changes in volatility when $\psi = 1$, as already seen in Chacko and Viceira (2005) . The results obtained in Hsuku (2007) further show that optimal consumptionwealth ratio is a function of stochastic volatility when $\psi \neq 1$: in particular it is an increasing function for investors whose EIS is smaller than one, while it is a decreasing function for investors whose EIS is larger than one. Finally, the analyses confirm the conclusions of Liu and Pan (2003) in the case of the power utility, regarding the role of derivatives: derivative securities are a significant tool for expanding investors' dimension of risk-and-return tradeoffs, being a vehicle for the additional risk factor of stochastic volatility in the stock market.

4 Conclusion

In this paper, a review of the literature on dynamic allocation has been proposed. In particular, some of the most popular models for the investor's choice strategy were analyzed, considering different preferences for the investor and market settings. It is observed how the use of affine models to model stochastic volatility leads to optimal weights that do not depend on the instantaneous volatility nor on the long-run mean level contradicting the foundations of the Markowitz portfolio theory, which theorize and demonstrates a relationship of inverse proportionality between the invested wealth and the synthetic indicators of variability.

Further contributions could be extended along two different lines: at first, the problem of optimal allocation in a complete market could be faced by considering the co-precision (the inverse of volatility) in a setting where there are discontinuities in both price and volatility to understand the impact of jumps and derivatives on the optimal portfolio in presence of precision when the investor has a power utility; at second the optimal allocation problem can be solved considering non-separable preferences for the investor (SDU utility function) and discontinuities in both price and volatility, analyzing both cases of complete and incomplete market.

References

Ait-Sahalia Y., Cacho-Diaz J., and Hurd T.R. (2009), Portfolio choice with jumps: a closed-form solution, *The Annals of Applied Probability*, 10(2), 556-584.

Branger N., Schlag C. and Schneider E. (2007), Optimal Portfolios when volatility can jump, *Journal of Banking and Finance*, 32(6), 10871097.

Buraschi A., Porchia P. and Trojani F. (2010), Correlation risk and optimal portfolio choice, *The Journal of Finance*, 65(1), 393-420.

Carr P., Jin X. and Madan D.B. (2001), Optimal investment in derivative securities, *Finance and Stochastics*, 4, 33-59.

Chacko G. and Viceira L.M. (2005), Dynamic Consumption and Portfolio Choice with Stochastic Volatility in Incomplete Market, *Review of Financial Studies*, 18(4),13691402.

Cheng Y. and Escobar-Anel M. (2021), Robust portfolio choice under the 4/2 stochastic volatility model, *IMA Journal of Management Mathematics*, 10.

Cox J.C. and Huang C.F. (1989), Optimal consumption and portfolio policies when asset prices follow a diffusion process, *Journal of Economic Theory*, 49, 33-83.

Das S. R. and Uppal R. (2004), Systemic Risk and International Portfolio Choice, *Journal of Finance*, 59 (6), 28092834.

Duffie D. and Epstein L.G. (1992), Stochastic differential utility, *Econometrica*, 60(2), 353-294.

Duffie D., Pan J. and Singleton K. (2000), Transform analysis and asset pricing for affine jump-diffusions, *Econometrica*, 16(68), 1343-1376.

Epstein L.G. and Zin S.E. (1989), Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework, *Econometrica*, 57(4), 937-969.

Escobar M., Ferrando S. and Rubtsov A. (2015), Robust portfolio choice with derivative trading under stochastic volatility, *Journal of Banking &Finance*, 61(2015), 142-157.

Faria G. and Correia-da-Silva J. (2016), Is stochastic volatility relevant for dynamic portfolio choice under ambiguity?, *The European Journal of Finance*, 22(7), 601-626.

Fisher M. and Gilles C. (1999), Consumption and Asset Prices with Homothetic Recursive Preferences, *Federal Reserve Bank of Atlanta*, 22(7), Working Paper 99-17.

Fleming W.H. and Hernandez-Hernandez D. (2003), An optimal consumption model with stochastic volatility, *Finance and Stochastics*, 7, 245-262.

Hansen L. P. and Singleton K. J. (1982), Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica*, 50(5), 1269-1286.

Haugh M. and Lo A. (2001), Asset allocation and derivatives, *Quantitative Finance*, 1, 45-72.

Hsuku (2007), Dynamic consumption and asset allocation with derivative securities, *Quantitative Finance*, 7(2), 137-149.

Ilhan A. and Sicar R. (2005), Optimal investment with derivative securities, *Finance and Stochastics*, 9, 585-595.

Jin X. and Zhang Allen X. (2012), Decomposition of optimal portfolio weight in a jump-diffusion model and its applications, *The Review of Financial Studies*, 25(9), 2877-2919.

Jin X., Luo D. and Zeng X. (2021), Tail risk and robust portfolio decisions, *Management Science*, 67(5), 3254-3275.

Karatzas I., Lehoczky J.P. and Shreve S. (1987), Optimal portfolio and consumption decisions for a small investor on a finite horizon, *SIAM J. Control Optimization*, 25, 1557-1586.

Kraft (2005), Optimal portfolios and Heston's stochastic volatility model: an explicit solution for power utility, *Quantitative Finance*, 5(3), 303-313.

Kraft H., Seifried F.T. and Steffensen M. (2013), Consumption-portfolio optimization with recursive utility in incomplete markets, *Finance and Stochastics*, 17, 161-196.

Kraft H., Seifried F.T. and Steffensen M. (2016), Optimal consumption and investment with Epstein-Zin recursive utility, *Finance and Stochastics*, 21, 187-226.

Liu J. and Pan J. (2003), Dynamic derivative strategies, *Journal of Financial Economics*, 69, 401-430.

Liu J., Longstaff A. and Pan J. (2003), Dynamic Asset Allocation with Event Risk, *The Journal of Financial Economics*, 58(1), 231-259.

Liu J. (2007), Portfolio selection in stochastic environments, *The Review of Financial Studies*, 20(1), 1-39.

Markowitz H. (1952), Portfolio Selection, *Journal of Finance*, 7, 77-91.

Merton R.C. (1971), Optimum consumption and portfolio rules in a continuous-time model, *Journal of Economic Theory*, 3(4), 373-413.

Oliva I. and Ren R. (2018), Optimal portfolio allocation with volatility and co-jump risk that Markowitz would like, *Journal of Economics Dynamics and Control*, 94, 242-256.

Pharm H. (2022), Smooth solutions to optimal investment models with stochastic volatilities and portfolio constraints, *Applied Mathematics and Optimization*, 46, 55-78.

Pharm H. (2007), Optimization methods in portfolio management and option hedging,*3rd cycle. Hanoi(Vietnam)*, 27, cel-00392212f.

Plisca R.S (1986), A stochastic calculus model of continuous trading: Optimal portfolios,*Math. Operations Res*, 11, 371-382.

Schroder M. and Skiadas C. (1999), Optimal Consumption and Portfolio Selection with Stochastic Differential Utility,*Journal of Economic Theory*, 89, 68-126.

Schroder M. and Skiadas C. (2003), Optimal Lifetime Consumption-Portfolio Strategies under Trading Constraints and generalized Recursive Preferences, *Stochastic Processes and their Applications*, 108, 155-202.

Von Neuman J. and Morgenstern O. (1947), Theory of games and economic behavior, *Princeton University Press.*

Xing H. (2017), Consumptioninvestment optimization with EpsteinZin utility in incomplete markets,*Finance Stochastics*, 21, 227262.

Zhou C., Wu C. and Wang Y. (2019), Dynamic portfolio allocation with time-varying jump risk, *Journal of Empirical Finance*, 50, 113-124.

SAPIENZA
UNIVERSITÀ EDITRICE

ANNALI DEL DIPARTIMENTO DI METODI E MODELLI
PER L'ECONOMIA IL TERRITORIO E LA FINANZA
2022 | pp. 81-88

E-ISSN 2611-6634 | ISSN 2385-0825
DOI: 10.13133/2611-6634/1408

Research article

Barbara Brollo*

SOGGETTI, EFFETTI E PRATICHE URBANE DELLE POPOLAZIONI TEMPORANEE. IL CASO DI ROMA

Abstract

The object of the research are temporary populations in urban contexts. Various forms of temporary dwelling are analysed in the literature on urban studies, such as those related to tourism, study, work and lifestyle migration, but they are treated mainly separately. Despite the strong differences between these groups, I propose temporariness as a dimension that connects them, in a macro-category that can be defined as temporary populations.

Through quantitative and qualitative methods, I present data sources, numerical estimates, socio-economic characteristics, and spatial trends of this category, in the specific case study of Rome. These data and analytical methods are useful to discuss the impact of these groups on housing, economic and social dynamics and, more generally, on the city's development model.

Keywords: temporariness, touristification, studentification, gentrification, lifestyle migrants.

* Sapienza Università di Roma

1 Introduzione

La ricerca, di cui in questo articolo si presenta una sintesi, approfondisce il tema dell'abitare temporaneo nelle grandi città contemporanee, in particolar modo quelle del Sud Europa e Roma quale caso studio specifico.

Faccio riferimento a forme di mobilità volontaria, in cui la transitorietà è una scelta. Tratto soprattutto categorie quali i turisti, gli studenti fuori sede, i migranti temporanei per lavoro, i lifestyle migrants, ovvero gruppi che insistono su determinate aree urbane per pochi giorni, mesi o anni, senza l'intenzione o la possibilità di trasferirvisi definitivamente. Analizzo congiuntamente queste categorie attraverso il concetto di popolazioni temporanee. Oltre ad una ricerca sulle tipologie di soggetti coinvolti, mi propongo di analizzare il loro comportamento spaziale e le pratiche che derivano dall'abitare temporaneo, sia a livello di motivazioni e prospettive individuali, che, soprattutto, in termini di effetto complessivo sul sistema urbano.

Dopo aver definito concettualmente l'oggetto di ricerca, i miei obiettivi riguardano la quantificazione e la localizzazione di popolazioni temporanee e una loro descrizione in termini di caratteri sociodemografici ed economici, nello specifico caso studio di Roma. Il proposito complessivo è quello di comprendere in che modo, e con quali conseguenze, una crescente porzione di varie città si orienta a un flusso di presenze transitorie. Gli strumenti che permettono il raggiungimento di questi obiettivi sono la ricognizione delle possibili fonti quantitative, l'elaborazione dei dati raccolti per ottenere delle stime numeriche e metodologie qualitative per approfondire gli aspetti più interpretativi.

2 Definizione e rilevanza del tema

La dicitura "popolazioni temporanee" non è particolarmente diffusa. I gruppi da cui è composta questa categoria sono oggetto di molte ricerche, soprattutto nell'ambito degli studi urbani, ma non vengono considerati congiuntamente, bensì più in forma frammentaria. I principali filoni di studio sul tema sono la turistificazione (Gotham, 2005; Nilsson, 2020; Palomeque, 2013), la studentificazione (Smith, 2002; Allison, 2006), la migrazione temporanea per lavoro (Khoo et al., 2008; Rahman, 2009; Mendoza et al., 2016), la lifestyle migration (Benson e O'Reilly, 2016) e le diverse sfumature del processo di gentrificazione associate a questi fenomeni (Gotham, 2005; Smith e Holt, 2007; Jover e Díaz-Parra, 2020; Sigler e Wachsmuth, 2020). Le popolazioni trattate dai diversi studi sono distinte principalmente sulla base delle motivazioni che le portano a muoversi in un determinato luogo, come il turismo, lo studio, il lavoro o la ricerca di esperienze. In questa ricerca propongo la categoria concettuale di popolazioni temporanee quale termine sotto il quale ricomprendere l'insieme di questi gruppi. Le popolazioni temporanee, seppur comprendendo sottogruppi molto eterogenei tra loro, hanno delle caratteristiche in comune che le rendono un insieme che è interessante da studiare, soprattutto in alterità e complementarità rispetto alla popolazione residente. La letteratura dimostra le particolarità di questi gruppi sotto diversi aspetti, per molti versi comuni, quali la giovane età (López-Gay et al., 2020), il potere acquisitivo tendenzialmente superiore a quello dei residenti dei quartieri in cui si insediano (López-Gay et al., 2020), i comportamenti volti al consumo (Bell e Ward, 2000), le loro particolari preferenze localizzative e di tipologia del domicilio (Garmendia et al., 2012), l'uso dello spazio pubblico (Nuvolati, 2003), la partecipazione nella politica locale (Martinotti, 1999), gli effetti economici in termini di occupazione e produzione di valore aggiunto (Cañada, 2018; Cheer, 2018).

La distinzione tra popolazione residente e non residente è rilevante per la comprensione di fenomeni urbani e dell'equilibrio a cui dà luogo (Martinotti, 1993), ad esempio in termini di composizione sociale, pratiche dell'abitare e prospettive di sviluppo. È interessante analizzare il rapporto tra residenti e questi sottogruppi di popolazione, anche in termini di possibili conflitti, ad esempio relativi all'uso e al finanziamento dei servizi e degli spazi pubblici (Nuvolati, 2003), e localizzativi, data la preferenza localizzativa di questi gruppi e l'effetto di espulsione, diretta o indiretta (Marcuse, 1985), che possono determinare in certe aree. Sul piano degli alloggi gli effetti di una competizione tra usi a lungo termine e usi temporanei si fanno sentire in termini di aumento dei prezzi di vendita e locazione, fino a una sostanziale indisponibilità di alloggi, soprattutto in affitto, per le finanze medie della popolazione residente, come dimostrano alcune ricerche empiriche (Schafer e Hirsch, 2017; Wachsmuth et al., 2017). Questa preoccupazione è rilevante soprattutto considerando un contesto economico e legislativo come quello del Sud Europa contemporaneo, che conta con una prevalenza di proprietà privata delle abitazioni, quindi una strutturale scarsità di alloggi in locazione, e liberalizzazione nel regime degli affitti (Siatitsa e Annunziata, 2017).

Come fanno notare Jover e Díaz-Parra in una delle loro analisi sul tema (2020), la gentrificazione intesa nel senso più classico del termine porta a una nuova comunità, a costo di dissolvere la precedente, ma almeno c'è una nuova collettività. Nel caso di sostituzione a opera di popolazioni temporanee, soprattutto quelle che si soffermano meno, ovvero i turisti, il cambio sociale porta all'assenza di una nuova comunità, a una sostituzione della residenzialità in sé con una presenza continua ma di soggetti sempre diversi. Come riportato da uno studio sull'economia turistica a Venezia, sembra che ci sia una preferenza delle città per la figura del "cittadino turista [che] non appartiene al luogo, ma vi partecipa essenzialmente con il consumo" (Ingersoll, 2004, p. 41 citato in Salerno, 2018, p. 487) piuttosto che il sostegno alla residenzialità di cittadini nel verso senso del termine.

3 Roma come caso di studio

Il caso di Roma è particolarmente interessante. Per quanto riguarda il turismo è la città con più presenze in Italia e tra le prime in Europa (ISTAT, 2020). Inoltre, essendo una capitale, è sede di ministeri e organizzazioni internazionali, ospita diverse università pubbliche e private, offre un patrimonio storico, artistico e architettonico molto conosciuto ed attrattivo. È inoltre capitale del cattolicesimo, data la presenza del Vaticano. Presenta quindi diversi motivi di attrazione. A questo aspetto si affianca il fatto che negli ultimi quarant'anni la parte della città compatta di Roma, composta dal centro storico e dai cosiddetti quartieri storici, abbia perso una popolazione pari a 800.000 persone, sia per effetto del calo demografico naturale che per i trasferimenti di residenza in aree più periferiche o in altri comuni, soprattutto quelli confinanti (Crisci, 2010). Se da una parte si elimina l'uso residenziale, a favore di attività terziarie, dall'altra si sostituiscono i residenti di lungo termine con abitanti temporanei.

Tra i principali problemi di amministrazione e gestione della città ci sono l'estensione abnorme del territorio comunale e la dipendenza dalle finanze pubbliche. Altra tendenza, collegata alla precedente, è un'economia a basso valore aggiunto. Nella mancanza di una struttura produttiva e innovativa, schiacciata tra assenza di investimenti pubblici e internazionali, Roma è una città in sofferenza. Nonostante la crisi di vari settori, uno dei meccanismi che continuano a funzionare è quello legato alla "rendita simbolica" (Tocci, 2020), cioè l'estrazione di valore dall'enorme, unica, ricchezza e varietà di reperti storici, manufatti artistici, ma anche di immaginari, sia quelli legati alla grandezza imperiale del passato, che quelli più contemporanei, veicolati da cinema e televisione. Nonostante abbia un forte, quasi naturale, potere di attrazione, Roma non riesce a trasformare questo potenziale in input per un reale sviluppo diffuso (Roma Ricerca Roma, 2021). La questione dell'estrazione della rendita, dal patrimonio fisico e abitativo, quale meccanismo privilegiato di produzione di valore, è a Roma una questione profondamente radicata nel suo regime urbano (D'Albergo e Moini, 2015). La logica della rendita, però, è quella di un'economia estrattiva e speculativa, che rischia di distruggere la ricchezza su cui prospera (Roma Ricerca Roma, 2021) e debilita le possibilità di uno sviluppo locale diffuso e condiviso. In questo meccanismo l'attrazione di popolazioni temporanee funge da cinghia di trasmissione tra figure in grado di alimentare il sistema di estrazione della rendita e luoghi che dipendono da questo meccanismo per la propria crescita economica. In questo quadro la velocità del ricambio di queste persone, quindi una permanenza sempre più transitoria, è considerabile quasi come un fattore che rende possibile il funzionamento del sistema.

4 Fonti e metodi per l'analisi quantitativa di popolazioni temporanee

La ricerca prende le mosse da uno studio della letteratura, italiana ed internazionale, riguardo le fonti, i dati e le metodologie di misurazione e descrizione di popolazioni temporanee. L'individuazione e quantificazione di questa macro-categoria è un compito arduo data la non unanime definizione e, soprattutto, per la loro natura che, dato il carattere transitorio, le rende difficilmente "catturabili". Date queste caratteristiche l'esplorazione di diverse tipologie di registri, da quelli più diffusi ed ufficiali a rilevazioni più specifiche, e tecniche di misurazione indirette, attraverso fonti non convenzionali, permettono di avere una panoramica su diverse fonti e metodologie di analisi.

È possibile sviluppare un'analisi sui dati disponibili nelle statistiche ufficiali, per quantificare, seppur approssimativamente, le popolazioni temporanee di Roma. Considerando congiuntamente i dati relativi agli arrivi per turismo, le anagrafi universitarie e le rilevazioni sulle forze lavoro si ha un quadro complessivo non solo della quantità di presenze temporanee in città ma anche di alcune caratteristiche socioeconomiche dei soggetti che compongono questi flussi.

Ci sono poi altre fonti, non appartenenti alle statistiche ufficiali, che permettono un ulteriore approfondimento della questione. La principale qualità di questi dati è che sono geo-riferibili, danno quindi una localizzazione a scala sub-comunale dei luoghi di pernottamento di questo macro-gruppo, per poter meglio analizzare l'effettiva incidenza nella sfera dell'abitare. Le principali criticità di questa tipologia di fonti riguardano la loro disponibilità e la frammentarietà dei fenomeni che catturano, in quanto legate specificamente ad alcune sottocategorie, come il turismo o lo studio. Alcuni dati geo-riferibili particolarmente interessanti, per il livello di precisione territoriale e per il grande potenziale informativo, sono stati resi disponibili dalla piattaforma Facebook in occasione delle straordinarie misure di limitazione della mobilità dovute alla pandemia da COVID-19. Questo evento, seppur nella drammaticità del contesto, può essere visto come un esperimento naturale, utile ad identificare le aree che più si spopolano a causa della volatilità di questi soggetti, che possono contare su altri luoghi dove vivere. Il dato proposto dalla piattaforma considera il calo percentuale di utenti presenti rispetto allo stesso momento nei tre mesi che precedono il giorno di riferimento. Rapportando questo dato alla popolazione residente è possibile ottenere una stima della popolazione fluttuante, quale gruppo presente nei mesi prima delle misure di limitazione degli spostamenti.

I dati analizzati dimostrano che la concentrazione e la centralità sono dimensioni interconnesse e definitorie della localizzazione di popolazioni temporanee, in linea con altri casi studio di grandi città europee (Maitland e Newman, 2008; Malet-Calvo et al., 2017; Novy, 2018).

5 Una ricerca ad hoc sulle popolazioni temporanee a Roma

Per esplorare la tematica in maniera più approfondita si è svolta una ricerca ad hoc, attraverso un questionario, distribuito tra maggio e giugno 2021 e delle interviste, condotte tra settembre e novembre 2021. Grazie alla comunicazione diretta con i soggetti implicati, questi strumenti di ricerca permettono di ricavare informazioni altrimenti inaccessibili, legate alle motivazioni, percezioni, immaginari, aspettative e pratiche abitative.

Il questionario è rivolto a persone che abitano a Roma da qualche mese o pochi anni, o che hanno abitato a Roma per un tempo limitato. È stato tradotto in inglese e spagnolo per facilitare la comprensione e compilazione da parte di persone di diverse provenienze. È stato creato e ospitato nella piattaforma Google Form e diffuso prevalentemente online, tramite piattaforme social, ma anche attraverso codici QR in volantini affissi in punti frequentati della città e con il passaparola. Le tematiche affrontate riguardano il motivo prevalente che ha portato a trascorrere un periodo a Roma e i fattori di attrazione per la città, questioni relative alla temporalità e temporaneità (abituale frequenza di periodi fuori città, se la si è lasciata durante il lockdown), pratiche abitative e intenzioni per il futuro.

Sono state raccolte 151 risposte valide. In seguito, ho svolto 15 interviste semistrutturate con alcuni dei rispondenti al questionario, selezionati sulla base della massima rappresentatività dei profili emersi. Non conoscendo l'universo di riferimento il campione dei rispondenti non si può considerare rappresentativo dell'intera popolazione di riferimento, ma è comunque un buon inizio per un'indagine esplorativa.

Dall'analisi delle risposte emergono dati interessanti. Si conta una grande varietà di origini geografiche, con ben 55 regioni italiane e 10 stati di provenienza e, nonostante la giovane età media dei rispondenti, attorno ai 30 anni, ben un quarto di essi ha già vissuto in altre città, oltre a quella di origine e Roma. Si conferma quindi una forte tendenza alla mobilità, nazionale ed internazionale. Un altro dato emerso, che conferma le ipotesi iniziali, è quello di un elevato livello di istruzione: l'87% dei rispondenti ha un titolo di studio equivalente o superiore alla laurea triennale, a fronte di una media nazionale del 20% di laureati. Riguardo le motivazioni che hanno portato allo spostamento, lo studio è quella prevalente (44%) ma è interessante notare come la dimensione dell'esperienza di vita sia un elemento trasversale, indicato come dimensione rilevante dalla maggioranza degli individui sottoposti alle interviste di approfondimento. Per la maggior parte dei rispondenti la temporaneità è un aspetto che chiaramente caratterizza la loro permanenza, in quanto non c'è un'intenzione definita di permanere a lungo in città, infatti il 31% dichiara di volersi fermare a Roma per un breve periodo, a fronte del 22% che sa di voler restare a lungo. Interessante anche che la maggior parte dei rispondenti (39%) dichiara di non sapere se resterà o no a lungo in città. Le tipologie di alloggio più frequenti e altri aspetti relativi alle pratiche abitative, come la localizzazione del domicilio e la frequenza e direzione della mobilità residenziale, presentano dinamiche particolari e riconoscibili, a conferma della coesione interna di questo gruppo rispetto alla popolazione residente.

6 Discussione e conclusioni

La temporaneità si rivela come un'utile categoria esplicativa di diversi flussi di mobilità che caratterizzano la contemporaneità. La crescente rilevanza di questo aspetto si manifesta in diverse forme; tra queste le pratiche relative all'abitare sono particolarmente rilevanti in quanto connettono una dimensione più intima e individuale con risvolti più ampi, che incidono sul sistema urbano.

Considerando l'impatto delle popolazioni temporanee non può mancare una riflessione sulle politiche urbane, spesso improntate all'attrazione di questi gruppi. Oltre alla dimensione del discorso politico e mediatico è interessante osservare i concreti investimenti, soprattutto in termini di grandi opere e di rigenerazione urbana, che sostengono questo processo di posizionamento delle città in termini di attrazione turistica e competitività per il talento. Sarebbe necessario analizzare il reale contributo di questi gruppi in termini di crescita e sviluppo e di sostenibilità sociale per gli abitanti delle città. La mobilità geografica, soprattutto se temporanea, può essere un elemento che amplifica le disuguaglianze preesistenti. La trasformazione di aree particolarmente attraversate da flussi di popolazione in rapido ricambio può essere interpretata come una dinamica di potere tra gruppi più abbienti, con minori vincoli legati a una dimensione locale, competenti nell'ambito digitale, in grado quindi di avere la meglio nella competizione per le risorse urbane (López-Gay et al., 2020).

Nello specifico caso di Roma, la città sembra trarre risorse più dal moto stesso delle persone, che si susseguono nel trascorrervi un periodo, che dalla creazione di valore prodotto dalla combinazione di capitale e lavoro. È quindi uno scambio quasi elettrico, meccanico, di esperienze emozionanti in cambio di consumo. La città a breve termine che ne consegue non è un sistema in grado di creare benessere diffuso ed è soggetta al cortocircuito.

Riferimenti bibliografici

Allison J. (2006), Over-Educated, Over-Exuberant and Over Here? The Impact of Students on Cities, *Planning, Practice & Research*, 21(1), 79-94. DOI: 10.1080/02697450600901541.

Bell M. e Ward G. (2000), Comparing Temporary Mobility with Permanent Migration, *Tourism Geographies*, 2(1), 87-107. DOI: 10.1111/1467-8470.00039.

Benson M. e O'Reilly K. (2016), From lifestyle migration to lifestyle in migration: Categories, concepts and ways of thinking, *Migration Studies*, 4(1), 20-37. DOI: 10.1111/j.1467-954X.2009.01864.x.

Cañada E. (2018), Too precarious to be inclusive? Hotel maid employment in Spain, *Tourism Geographies*, 20(4), 653–674. DOI: 10.1080/14616688.2018.1437765.

Cheer J. (2018), Geographies of marginalization: Encountering modern slavery in tourism, *Tourism Geographies*, 20(4), 728–732. DOI: 10.1080/14616688.2018.1477171.

Crisci M. (2010), *Italiani e stranieri nello spazio urbano. Dinamiche della popolazione di Roma*. Franco Angeli, Milano.

D'Albergo E. e Moini G. (2015), *Il regime dell'urbe: politica, economia e potere a Roma*. Carocci, Roma.

Garmendia M., Coronado J. M. e Ureña J. M. (2012), University Students Sharing Flats: When Studentification Be-comes Vertical, *Urban Studies*, 49(12). DOI: 10.1177/0042098011428176.

Gotham K. (2005), Tourism gentrification: The case of New Orleans' Vieux Carré (French Quarter), *Urban Studies*, 42(7), 1099–1121. DOI: 10.1080/14616688.2019.1586986.

Ingersoll R. (2004), *Sprawltown*. Meltemi, Roma.

ISTAT (2020), *Movimento turistico in Italia*, disponibile online https://www.istat.it/it/files/2020/12/REPORT_TURISMO_2020.pdf [14 marzo 2022].

Jover J. e Díaz-Parra I. (2020), Who is the city for? Overtourism, lifestyle migration and social sustainability, *Annals of Tourism Research*, 23(2), 261-283. DOI: 10.1080/14616688.2020.1713878.

Khoo S., Graeme H. e Mcdonald, P. (2008), Which Skilled Temporary Migrants Become Permanent Residents and Why? *The International Migration Review*, 42(1), 193–226. DOI: 10.1111/j.1747-7379.2007.00118.x.

López-Gay A., Cocola-Gant A. e Russo A. P. (2020), Urban tourism and population change: Gentrification in the age of mobilities, *Population, Space and Place*, 27(1), 1-17. DOI: 10.1002/psp.2380.

Maitland R. e Newman P. (2008), Visitor-host relationships: conviviality between visitors and host communities. In Hayllar B., Griffin T., Edwards D. (a cura di), *City spaces, tourist places: Urban tourism precincts.* Butterworth-Heinemann, Oxford.

Malet-Calvo D., Nofre J. e Geraldes M. (2017), The «Erasmus Corner»: Gentrification, emotions and place-making of a nightlife spot in Bairro Alto, *Leisure Studies*, 36(6), 778-792. DOI: 10.1080/02614367.2016.1271821.

Marcuse P. (1985), Gentrification, abandonment, and displacement: Connections, causes, and policy responses in New York City, *Journal of Urban and Contemporary Law*, 28, 195-240.

Martinotti G. (1993), *Metropoli, la nuova morfologia sociale della città.* Il Mulino, Bologna.

Martinotti G. (1999), A city for whom? Transients and public life in the second-generation metropolis. In Beauregard R., Body-Gendrot S. (a cura di), *The Urban Moment. Cosmopolitan Essays on the Late-20th-Century City*. Sage, London.

Mendoza C., Staniscia B. e Ortiz Guitart A. (2016), Migración y movilidad de las personas calificadas: nuevos enfoques teóricos, territorios y actors, *Geocritica*, XXI, 1-22. DOI:.1344/b3w.0.2016.26353.

Nilsson J. H. (2020), Conceptualizing and Contextualizing Over-tourism: the Dynamics of Accelerating Urban Tourism, *International Journal of Tourism Cities*, 6(4), 657-671. DOI 10.1108/IJTC-08-2019-0117.

Novy J. (2018), 'Destination' Berlin revisited. From (new) tourism towards a pentagon of mobility and place consumption, *Tourism Geographies*, 20(3), 418-442. DOI:10.1080/14616688.2017.1357142.

Nuvolati G. (2003), Resident and Non-resident Populations: Quality of Life, Mobility and Time Policies, *The Journal of Regional Analysis and Policy*, 33(2), 1-17. DOI: 10.22004/ag.econ.132256.

Palomeque F. P. (2013), Barcelona, de ciudad con turismo a ciudad turística. Notas sobre un proceso complejo e inacabado, *Documents d'Anàlisi Geogràfica*, 61(3), 483-506. DOI: 10.5565/rev/dag.33.

Rahman M. M. (2009), Temporary Migration and Changing Family Dynamics: Implications for Social Development, *Population, Place and Space*, 15(2), 161-174. DOI: 10.1002/psp.537.

ROMA RICERCA ROMA (2021), *Un manifesto per Roma. Il diritto a una città giusta. Percorsi per uscire dalla crisi del valore*, disponibile online https://www.ricerca-roma.it/wp-content/uploads/2021/04/Roma-Ricerca-Roma-Il-diritto-a-una-citta%CC%80-giusta.pdf [14 marzo 2022].

Salerno G. M. (2018), Estrattivismo contro il comune. Venezia e l'economia turistica, *ACME*, 17(2), 480-505.

Schäfer P. E e Hirsch J. (2017), Do urban tourism hotspots affect Berlin housing rents? *International Journal of Housing Markets and Analysis*, 10(2), 231–255. DOI: 10.1108/IJHMA-05-2016-0031.

Siatitsa D. e Annunziata S. (2017), Attempts to prevent displacement: housing policies in time of austerity in Athens and Rome. In Annunziata S., Lees L. (a cura di), *Anti-gentrification nelle città (Sud) Europee*. Urbanistica Tre, Roma.

Sigler T. e Wachsmuth D. (2020), New directions in transnational gentrification: Tourism-led, state-led and lifestyle-led urban transformations, *Urban Studies*, 57(15), 3190-3201. DOI: 10.1177/0042098020944041.

Smith D. (2002), Patterns and processes of studentification in Leeds, *The Regional Review*, 12, 14-16.

Smith D. e Holt L. (2007), Studentification and 'apprentice' gentrifiers within Britain's provincial towns and cities: extending the meaning of gentrification, *Environment and Planning A*, 39,142-161. DOI: 10.1068/a38476.

Tocci W. (2020), Roma come se: alla ricerca del futuro per la capitale. Donzelli, Roma.

Wachsmuth D., Kerrigan D., Chaney D. e Shillolo A. (2017), *Short-term cities: Airbnb's impact on Canadian housing markets*. McGill University, School of Urban Planning, Montreal, disponibile online https://upgo.lab.mcgill.ca/publication/short-term-cities/ [14 marzo 2022].

Ringraziamenti

Ringrazio il professor Filippo Celata per l'attenta supervisione e gli insegnamenti trasmessi. Ringrazio anche i miei compagni di dottorato per il supporto e gli stimoli.

Sapienza Università Editrice
ANNALI DEL DIPARTIMENTO DI METODI E MODELLI PER L'ECONOMIA IL TERRITORIO E LA FINANZA
2022 | pp. 89-105

E-ISSN 2611-6634 | ISSN 2385-0825
DOI: 10.13133/2611-6634/1389

Research article

Adriana Conti Puorger *

SOSTENIBILITÀ VERSO RECOVERY. UNA LETTURA DELLA NARRAZIONE DEL CAPITALE NATURALE

Abstract

The setting of the Ecological Economy has initiated new reflections on the issues of local development and has underlined the aspects of governance and relations promoted by local actors. The contribution, starting from the mitigation actions initiated in some glacial bodies of the Adamello-Presanella mountain range, reflects on the links between sustainability and recovery as they emerge from the discussions on ecosystem services and their enhancement. The reference framework for the theme of ecological transition and integration between economy and ecology is reconstructed by retracing the dynamics of global governance on the issues of sustainability and, in particular, biodiversity. The dynamics and concretization of all these aspects investigated at the local scale allows a reflection on the creation/appropriation of the value of natural capital and on local thinking about the meaning of sustainability.

Keywords: Glacier, natural capital, recovery.

---

* Sapienza University of Rome

1 Introduzione

Ci sono dei luoghi sentinella nei quali le criticità del cambiamento climatico sono particolarmente evidenti tanto da assumere il ruolo di indicatori di intensità di tali andamenti (EEA, 2017). I ghiacciai rientrano in questa categoria e, conseguentemente, sono luoghi sui quali indagare gli effetti innescati dal paradigma della sostenibilità sulla gestione dei temi della mitigazione e dell'adattamento. In questa traiettoria il contributo si propone di guardare alle relazioni che gli attori locali hanno posto in essere rispetto alla mitigazione delle riduzioni dei ghiacciai con attenzione ad alcune vedrette del gruppo dell'Adamello-Presanella. Dal 2008 parte di tale ghiacciaio, infatti, è coperto da teli geotessili, tecnica proattiva nell'azione di mitigazione che è svolta in molte aree alpine anche austriache e svizzere. Tale pratica quasi ventennale è stata oggetto di analisi e valutazioni che sostengono i parziali benefici della riduzione di assorbimento di energia solare da parte delle superfici trattate e, dunque, di salvaguardia della risorsa. Il termine risorsa in riferimento alla Natura introduce, come verrà esplicitato nel lavoro, il tema del capitale naturale e della sua valorizzazione per lo sviluppo locale, pratica della sostenibilità così come si è sviluppata nell'approccio dell'Ecological Economics. I concetti di servizi ecosistemici e di biodiversità (Costanza et al., 1997; Barrett, Farina, 2000; Brown, 2001), introdotti e discussi in tale interdisciplinare campo di studi, hanno il fine di fornire un ponte tra le discipline economiche e quelle ecologiche in una nuova prospettiva di gestione sostenibile. L'approccio socio-ecosistemico che ne è scaturito, infatti, ha contribuito a costruire schemi interpretativi che, con il contributo della valutazione dei servizi ecosistemici, permettono la commensurabilità tra le variabili economiche e quelle ambientali.

Interrogarsi sul significato di tali interfacce è un modo per indagare sul concetto di transizione, di scambio tra schemi interpretativi e passaggio di concetti che vanno, come nel caso in esame, dal patrimonio al capitale, dalla conservazione alla valorizzazione, dal valore intrinseco a quello strumentale.

Il ghiacciaio, dunque, sostanzia il capitale naturale e fornisce servizi ecosistemici di approvvigionamento, di regolazione e culturali secondo la schematizzazione del Millenium Ecosystem Assessment (MEA, 2005). É, infatti, funzionale per l'economia locale sia per la fornitura di acqua e di energia ma, particolarmente, è risorsa economica per l'industria turistica, da cui l'incipit per l'orientamento verso azioni di mitigazione. Il punto che vuole essere osservato è il ruolo dell'influenza del contesto socioculturale in cui viene a promuoversi l'azione di sostenibilità. Le modalità della relazione uomo-natura (e contermini) delle azioni proposte dagli attori locali possono essere inserite, infatti, in letture economiche rispetto a logiche utilitaristiche o solidali (Rossi, 2008). Quali conseguenze derivano dalla relazione ecologia/economia basata sulla lettura funzionale introdotta dalla Economia Ecologica? È rintracciabile un processo di transitività che induce, paradossalmente, ad attribuire all'unità ecosistemica un valore intrinseco grazie al suo valore strumentale?

La lettura funzionalista delle pratiche della sostenibilità, dunque, in che modo influisce e orienta il tema della produzione sociale della natura? Quali spazi, quali paesaggi e quali società scaturiscono da questa impostazione? Quesiti proposti nelle varie anime interpretative della Social Nature Geographies (Bonati, Tononi, Zanolin, 2021) che permettono di riflettere sulle dinamiche di creazione/appropriazione del valore del capitale naturale e sono state di guida per introdurre una critica alla narrazione della sostenibilità condotta in un determinato contesto.

Il contributo vuole evidenziare l'interpretazione della sostenibilità attuata in un sistema ibridato tra lo stereotipo del modello capitalista e quello delle economie costruite intorno allo sviluppo locale. Ripercorrere le relazioni trans-scalari che sono all'origine del discorso sui servizi ecosistemici e sul capitale naturale, è quanto presentato nella prima parte del lavoro che, successivamente, si sofferma sui principali adeguamenti del contesto italiano a tali dinamiche. Il caso studio del Presena, parte applicativa del lavoro, presenta la scala locale che incorpora i temi dell'interdipendenza tra sviluppo e sfruttamento delle risorse. Infine, l'importanza delle dinamiche subglobali nell'influenzare il funzionamento del sistema Terra così come il rapporto tra scienza e politica suscitano l'interesse per delle riflessini sul senso delle azioni intraprese e progettate, cercando le connessioni tra gli imperativi globali, la compartecipazione nazionale e le dinamiche alla scala locale, temi delle osservazioni conclusive.

2 Ecologia Economica e capitale naturale

È noto il ruolo assunto dal tema della sostenibilità dal 1987 con la presentazione del rapporto «Our common future» della Commissione mondiale su Ambiente e Sviluppo (World Commission on Environment and Development, WCED). Quella fu l'occasione della formulazione di una linea guida ancora oggi valida che, attraverso la Conferenza delle Nazioni Unite su ambiente e sviluppo di Rio nel 1992, orientò la governance globale su tali temi con l'avvio di tre importanti Convenzioni, che presentano percorsi differenti ma convergenti rispetto ai temi ambientali. Il riferimento è alla Convenzione sui Cambiamenti Climatici, corroborata dalle valutazioni scientifiche di risalto mondiale dell'Intergovernmental Panel on Climate Change (IPCC,1988), da cui il protocollo di Kyoto e il successivo Accordo di Parigi con gli impegni fino al 2030; la Convenzione sulla lotta alla desertificazione, e i relativi Piani d'Azione Nazionali, e la Convenzione sulla Biodiversità (CBD), entrata in vigore nel 1993, di cui si tratterà brevemente rispetto ai principali impulsi e progressi che da questa provengono poiché la ricongiungono, come per le altre convenzioni, sia ad Agenda 2030 e, in particolare, al discorso sul capitale naturale.

L'aspetto innovativo legato alla tematica della biodiversità è l'attenzione per la dimensione ecosistemica che integra gli elementi biotici e abiotici presenti e necessari a determinarla. È, dunque, una scala per le azioni di tutela e conservazione che scavalca la singola specie o quella di particolari aree protette perché sottolinea i legami che esistono e insistono sulle dinamiche ecosistemiche di cui la società è una componente (CDB, 1992). Un primo processo transitivo nella logica della costruzione di schemi interpretativi della sostenibilità che a questo punto si percepisce è che, come negli ecosistemi, si ha un'integrazione dinamica e sinergica tra le componenti, così bisogna orientare le azioni politiche a integrare gli aspetti sociali ed economici per conseguire il risultato del benessere. Il secondo processo transitivo, quasi come feedback del precedente, si riscontra con il tema dei servizi ecosistemici, terziario naturale, che può essere contabilizzato dall'economia e, dunque, fornire la piattaforma integrativa tra economia ed ecologia. Orientamento, questo, che trova la sua sintesi nel lavoro Millennium Ecosystem Assessment (MEA, 2005). È possibile, su questa scia, articolare maggiormente il quadro di riferimento tra servizi ecosistemici e capitale naturale. Un importante ruolo in tal senso, che si aggiunge al lavoro sul MEA, si trova nel testo Handbook of National Accounting: Integrated Environmental and Economic Accounting (ONU, EC, FMI, OCSE, WB, 2003)[1], in cui si sviluppano gli aspetti dell'integrazione della contabilità ambientale in quella economica, dunque della commutabilità. In tale sede viene esplicitato che il capitale naturale è connaturale ai temi della biodiversità e che comprende tre categorie principali: stock di risorse naturali, territorio ed ecosistemi. Tutti sono considerati essenziali per la sostenibilità a lungo termine dello sviluppo per la loro fornitura di "funzioni" all'economia, così come all'umanità. Nel 2007 a seguito della riunione dei ministri dell'Ambiente a Potsdam che ribadivano la necessità di un'analisi sul significato economico della biodiversità, nasce, inoltre, l'iniziativa TEEB (The Economics of Ecosystems & Biodiversity) che avvia una sistematizzazione della valutazione del capitale naturale.

I concetti di servizi ecosistemici e di biodiversità, come anticipato, sono stati un modo per costruire interfacce atte a realizzare dei modelli di economia che interiorizzassero i temi della valutazione delle esternalità ambientali e della loro contabilità in modo da esplicitare le relazioni tra sostenibilità ambientale e sociale (Costanza et al., 1997). L'unità di analisi ecosistemica mette in relazione ecologia ed economia: la salvaguardia della integrità di tale interazione diviene la misura della capacità di preservare la stabilità ecologica e, dunque, quella economica. Il sistema ecologico, dunque, è un archetipo di sostenibilità e adeguare le metriche dello sviluppo rispetto ai legami complessi come quelli che si attuano nell'unità ecosistemica è funzionale per l'orientamento verso nuovi modelli. Il buon funzionamento dell'ecosistema sostiene i processi di sviluppo. Scorrendo il percorso in cui è maturata l'attenzione alla gestione della biodiversità si evidenziano, infine, i tasselli del quadro in cui si è sviluppata la modalità di misurazione del capitale naturale.

2.1 Gli obiettivi e le misure

Fa parte di questa ricostruzione la dichiarazione ONU sugli Obiettivi del Millennio (MDG) negli anni 2000 con l'impegno a garantire la sostenibilità ambientale declinandola in quattro sotto temi e undici indicatori tra i quali si cita, per l'attinenza con il presente lavoro, quello sulla riduzione del processo di annullamento della biodiversità per raggiungere, nel 2010, un contenimento significativo del fenomeno. Anche nel secondo Summit mondiale sullo sviluppo sostenibile nel 2002 a Johannesburg la governance globale sui temi della biodiversità si formalizza ribadendo l'impegno per l'obiettivo 2010 di cui la CBD diviene l'organo responsabile per il suo raggiungimento.

---

[1] Per una breve ricostruzione dei principali riferimenti sul tema della valutazione dei servizi ecosistemici, a partire dalla Commissione Mondiale sull'Ambiente e lo Sviluppo (1983-1987) e con la conseguente Agenda 21 (1992), si ricorda nel 1993 il testo Handbook of National Accounting: Integrated Environmental and Economic Accounting (SEEA, System of Environmental Economic Accounting, 1993), aggiornato in Handbook of National Accounting: Intergrated Environmental and Economic Accounting (SEEA-2003). Lavori che sono continuati con il Committee of Experts on Environmental-Economic Accounting nel 2005. Più recenti il lavoro: Mainstreaming the Economics of Nature: A Synthesis of the Approach, Conclusions and Recommendations of TEEB (The Economics of Ecosystems and Biodiversity) del 2010. Sempre alla scala globale UNEP/UNSD/CBD project on Advancing Natural Capital Accounting, SEEA Experimental Ecosystem Accounting: Technical Recommendations, 2017, e l'ultimo riferimento in campo interanzionale: Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science- Policy Platform on Biodiversity and Ecosystem Services (IPBES, 2019)

Successivamente, nel 2010 è stato concordato anche un piano strategico decennale volto a contrastare la perdita di biodiversità definendo venti obiettivi, noti come gli obiettivi di Aichi. Piano strategico che, però, non ha raggiunto nessuno dei target che si era prefissato (Global Biodiversity Outlook 5).

Altro tema importante della biodiversità, già presente nell'impostazione del piano strategico della Convenzione del 1992, è l'impegno a forme di tutela e cooperazione riguardo agli aspetti di accesso e uso delle risorse genetiche. Il processo avviato approderà a due protocolli quello di Cartagena (2000) sulla biosicurezza, che disciplina i movimenti di organismi viventi modificati dalla biotecnologia moderna da un paese all'altro, e quello di Nagoya (2010), relativo all'accesso e alla condivisione dei benefici derivanti dall'uso delle risorse genetiche. Nel frattempo, con il già ricordato rapporto MEA, assume rilevanza anche il tema dei pagamenti dei servizi ecosistemici, rivolto allo sviluppo di modelli capaci di cogliere in chiave economica le relazioni uomo ambiente. La biodiversità è, dunque, di sostentamento anche per gli aspetti dell'innovazione della tecnologia dell'economia e della finanza, considerando i fondi proposti e istituiti per le strategie in esame, che, si ricorda, erano partite come lotta alla povertà (Our common future, 1987). In questo vortice dei Summit mondiali, delle Convenzioni, dei protocolli, dei piani strategici e dalle loro tempistiche prende corpo la governance ambientale multilivello con un interscambio tra le scale macroregionali e nazionali e, via via, alle scale regionali e locali. Anche Agenda 2030, che rilancia nel 2015 i diciassette obiettivi di sostenibilità, è uno dei momenti di raccordo e di nuova propulsività nel dibattito mondiale della sostenibilità. Il lessico della sostenibilità con le emergenze e le urgenze suffragate dalle tante inconfutabili valutazioni scientifiche, come il lavoro sulla lettura dei limiti planetari proposti da Rockström (Rockström, et al., 2009) è fortemente correlato e accolto nella prospettiva della transizione ecologica aggiornata alla scala globale con gli obiettivi di sviluppo sostenibile approvati, appunto, nell'Agenda 2030 (2015), e gli impegni economici assunti.

Riallacciando l'esame delle dinamiche in seno alle concatenazioni trans-scalari della governance ambientale, si propone un breve e non esaustivo elenco dei principali atti a livello europeo sia rispetto agli argomenti propri degli aspetti ambientali sia di quelli dello sviluppo, dalla direttiva Habitat (1992); alla Strategia Pan Europea sulla Diversità Biologica e Paesistica (1996); successivamente la Convenzione del Paesaggio (2000), temi che riecheggiano nelle Strategie di Lisbona/Goteborg; Europa 2020 e, ad oggi, il Green deal europeo, parte integrante della strategia della Commissione Europea per l'attuazione dell'Agenda 2030. Ripercorrendo i sentieri europei si ricorda, in particolare, che all'interno della Strategia Europa 2020 la Commissione ha presentato, nel gennaio 2011, la settima iniziativa-faro: Un'Europa efficiente sotto il profilo delle risorse. Occasione nella quale sono state indicate le principali politiche e strategie in materia. Ne rappresenta parte integrante la Strategia sulla Biodiversità[2] che si inserisce nel percorso della volontà di arrestare la perdita di biodiversità (entro il 2020) e il degrado degli ecosistemi (entro il 2050) nell'Unione europea (UE) [COM (2011) 244 def.]; [COM (2013) 249 final].

---

[2] Si ricordano le direttive principali in tema di biodiversità: Direttiva 2009/147/CE concernente la conservazione degli uccelli selvatici; Direttiva del consiglio 92/43/CEE relativa alla conservazione degli habitat naturali e seminaturali e della flora e della fauna selvatiche. Direttive che hanno permesso la costruzione Natura 2000, formando rete di zone di protezione speciale (ZPS) e siti di interesse comunitario (SIC)

2.2 Capitale Naturale Italia

A questo milieu aderisce anche l'Italia con la Strategia Nazionale per la Biodiversità (SNB, Ministero dell'Ambiente e della tutela del territorio e del mare MATTM, 2010), che esplicita l'orizzonte a cui si vuole tendere attraverso un approccio ecosistemico: «La biodiversità e i servizi ecosistemici, nostro capitale naturale, sono conservati, valutati e, per quanto possibile, ripristinati, per il loro valore intrinseco perché possano continuare a sostenere in modo durevole la prosperità economica e il benessere umano nonostante i profondi cambiamenti in atto a livello globale e locale» (SNB, 2010, p.13). Nel 2011, facendo seguito alla Strategia, si è avuta l'istituzione degli organismi per la sua implementazione: il Comitato Paritetico per la Biodiversità (a supporto della Conferenza Stato-Regioni) e l'Osservatorio Nazionale per la Biodiversità (che cura l'apporto scientifico multidisciplinare composto da rappresentanti degli Osservatori o Uffici regionali sulla biodiversità, delle principali Associazioni Scientifiche, del mondo accademico, dall'ISPRA e dalle Aree Protette).

Il coinvolgimento degli attori nazionali è anche indice di condivisione di visioni e di lessico in un processo relazionale tra le varie scale che filtra gli orientamenti alle scale sub statali.

La declinazione della strategia, nel rispetto della ripartizione delle competenze dettata dalla riforma del Titolo V, è attuata dallo Stato per la competenza legislativa esclusiva in materia di «Tutela dell'ambiente e degli ecosistemi» (Costituzione, art. 117, comma II, lett. S), mentre trasferisce alle Regioni e agli altri Enti locali specifiche competenze gestionali nei diversi settori. La dotazione di strumenti conoscitivi e gestionali è, a sua volta, articolata tramite il Network Nazionale per la Biodiversità, punto d'incontro multilivello e multidisciplinare sui nuovi imperativi dettati dalla sostenibilità.

Attraverso il network si sono raggiunti importanti obiettivi sul tema dei servizi ecosistemici e della loro analisi come emerge dal rapporto «Stato della Biodiversità in Italia – Contributo alla strategia nazionale per la biodiversità» (2015) con la classificazione e valutazione quantitativa dei servizi e delle corrispondenti misure per la conservazione e riqualificazione degli ecosistemi e la loro valorizzazione[3], realizzando quel ponte proprio dell'articolazione della Economia Ecologica tra tutela della Natura e sua valorizzazione.

---

[3] Le principali azioni svolte rispondono alla necessità di recepire gli impegni internazionali assunti. Le informazioni cartografiche e le banche dati di valenza nazionale sono confluite nel Sistema Ambiente 2010 che rappresenta il progetto di innovazione digitale nel settore della protezione della natura con riferimento alla Biodiversità ed alle Aree Naturali Protette ed è volto a integrare il Network Nazionale per la Biodiversità (NNB) e il GeoPortale Nazionale. Il primo risponde alla necessità di raccolta, coordinamento e produzione della conoscenza tematica secondo metodiche standardizzate e certificate (Decreto istitutivo NNB PNM-DEC-2012-0000267). Il Network è in grado di garantire l'interoperabilità con analoghe infrastrutture internazionali (LifeWatch, GBIF, etc.) e con il GeoPortale Nazionale, in coerenza con quanto previsto dalla Direttiva INSPIRE (D.Lgs. 32/2010).

È possibile riconoscere tale passaggio guardando, in particolare, alla categoria delle Aree Naturali Protette, soggetti attivi nella Strategia anche per il loro ruolo di rappresentanti delle ecoregioni[4] (Blasi et al, 2010). Le prime sono istituzioni che rappresentano i beni del demanio naturale e sono disciplinate dalla legge del 6 dicembre 1991 n. 394 nella quale si fa espresso riferimento al patrimonio naturale (art.1 primo comma) e alla sua conservazione. Le seconde, invece, sono unità territoriali ecologicamente omogenee oggetto delle azioni di valorizzazione, previste nelle strategie sull'ambiente.

Il delicato legame tra i termini patrimonio naturale e capitale naturale è, dunque, l'incorporea piattaforma del paradigma della sostenibilità che innesca la ricerca di un linguaggio comune per integrare, come già detto, la componente ambientale a quella economica, e dunque sociale. Il nesso tra il valore contabile delle risorse naturali e il valore ecologico dei servizi ecosistemici diviene necessario per la tutela, la gestione e valorizzazione del patrimonio naturale, situato nelle ecoregioni. L'identificazione e classificazione dei servizi ecosistemici, la schematizzazione della relazione tra tali servizi e il benessere umano, evidenzia come quest'ultimo dipenda dalla dinamica della biodiversità che è l'indicatore per valutare nel tempo la capacità della natura di svolgere i suoi compiti.

Il tema del valore della biodiversità è, dunque, il raccordo tra ecologia ed economia, collegando le dimensioni spaziali dei servizi ecosistemici, situati in specifici luoghi, di cui lo Stato garantisce la tutela, al processo decisionale per il loro uso nelle diverse scale regionali attente alla valorizzazione. Con l'aumento del grado di consapevolezza rispetto alla dipendenza umana dalle risorse naturali, infine, l'operatore pubblico sarà indotto a rafforzare gli impegni economici e finanziari per il riconoscimento del loro valore (CBD/COP/DEC/X/2 (2010)).

2.3 Ulteriori aspetti documentali e procedurali

In Italia il rapporto tra patrimonio e capitale naturale e, dunque, le nuove dinamiche della tutela, conservazione, valorizzazione e gestione, è rintracciabile nei documenti che vanno dal primo rapporto sulla biodiversità e sull'attuazione della SNB relativo al biennio 2011-2012; al secondo Rapporto sulla SNB 2011-2020 (in relazione al Piano Strategico per la biodiversità 2011-2020 compresi gli obiettivi di Aichi per la biodiversità) e, poi, il terzo rapporto 2015-2016 che segna il momento di passaggio, derivato dalla Legge n. 221 del 28 dicembre 2015, in materia di capitale naturale. Con la costituzione del Comitato per il capitale naturale, infatti, si ha il primo rapporto sullo Stato del Capitale Naturale in Italia 2017-2018, e, ad oggi, è stato pubblicato il IV Rapporto.

Tutto questo converge nel tema della grande opera pubblica di ripristino degli ecosistemi di cui il PNRR (Piano Nazionale di Ripresa e Resilienza, 2021) rappresenta l'aggiornamento di quanto costruito nelle convezioni già richiamate e da Agenda 2030, nonché anche da quanto auspicato nel settimo programma di azione ambientale, ribadito e arricchito nell'ottavo programma di azione ambientale 2021-2030 dell'Unione. Atti che prevedono importanti impegni economici di azioni per il clima, per l'adattamento ai cambiamenti climatici e per la biodiversità. È evidente, a questo punto, l'imponente

---

[4] Per una ricostruzione di tale areale si rimanda al lavoro di M.Bagliai e A.Pietta, Regione Sostenibile e Bioregione, Regioni e Regionalizzazione, (eds) Gavinelli D., M. Bolocan Goldstein, 2022,Pearson, Milano-Torino.

forza della finanza ambientale che è collegata al tema della valorizzazione dei servizi ecositemici.

In continuità con quanto esposto i ghiacciai sono delle unità ecosistemiche e forniscono tre particolari tipologie di servizi: di approvvigionamento, di regolazione e culturali secondo la classificazione dei servizi ecosistemici in Italia per tipologie territoriali (Giupponi, Galassi, Pettanella, 2009). Rispetto al territorio scelto come oggetto di studio i ghiacciai ricadono nella classificazione gerarchica delle ecoregioni della provincia del sistema alpino meridionale 11C sezione delle alpi Retiche, ed interessano l'Area protetta: Parco Regionale Adamello (Regione Lombardia) e Parco Naturale Adamello-Brenta (Provincia Autonoma di Trento). Per una indicazione di massima sul valore dei servizi ecosistemici estraibili dalle aree montane quali la fornitura di acqua, la regolazione del clima, la fornitura di servizi culturali si riporta quanto presentato dagli esperti del Comitato per il Capitale Naturale Italiano (CCN) nel convegno scientifico: Conoscere, Valutare, Conservare e Arricchire il Capitale Naturale in Italia. Il valore monetario relativo alla qualità degli habitat è pari a 13,5 Mld € di cui: 4,9 Mld € attribuibili alle foreste, 1,5 Mld € a prati e pascoli e 103 Mln € alle zone umide; il valore monetario del servizio di attività ricreativa outdoor vale circa 8,4 Mld € nel 2018; il valore monetario del servizio di approvvigionamento idrico vale circa 1,3 Mld € nel 2018 (Capriolo, 2020).

3 Gruppo montuoso Adamello-Presanella e criosfera

La criosfera indica quella parte della superficie terrestre che risulta ricoperta dai ghiacci, è elemento del sistema climatico e concorre alla determinazione del clima terrestre e delle sue variazioni. Tale componente è stata oggetto della pubblicazione Special Report On the Ocean and Cryosphere in a Changing Climate dell'IPCC (Intergovernmental Panel on Climate Change) nel 2019. Nello Special Report si presentano i risultati degli studi a scala internazionale nella misurazione delle dinamiche climatiche delle aree di alta montagna nonché delle indicazioni sugli effetti tra le azioni della società e i servizi di tali ecosistemi. A seguito del riscaldamento globale, infatti, i cambiamenti che si registrano in tali aree, ossia in quelle regioni montane che presentano ghiacciai, neve o permafrost quali caratteristiche importanti del paesaggio, impattano negativamente sui sistemi fisici, biologici e umani delle montagne e delle pianure circostanti evidenziando la fragilità di tali sistemi. I ghiacciai, in particolare, sono sensibili alle fluttuazioni climatiche e offrono una evidenza delle dinamiche in corso. Le loro misurazioni e la variabilità dei dati raccolti sono così dei validi indicatori climatici. I monitoraggi, a scala internazionale, prendono avvio con il primo catasto globale dei ghiacciai (escluse le calotte antartica e groenlandese), del World Glacier Inventory (WGI); progetto avviato alla fine degli anni Cinquanta e completato negli anni Ottanta, aggiornato e ampliato nel 1989 dal World Glacier Monitoring Service (WGMS). Le innovazioni nel campo della cartografia digitale trovano espressione nel Randolph Glacier Inventory (RGI) e nelle sue versioni via via aggiornate.

Le misurazioni dei ghiacciai dell'arco alpino hanno una più lunga tradizione, si può risalire alla carta della Svizzara di Aegidius Tschudi che nel 1538 rappresentò un ghiacciaio alpino con il toponimo «Der Gletscher», o nel 1774 «Atlas Tyrolensis» ad opera di Anich e Hueber, dove si rintraccia il più antico riferimento cartografico al toponimo Adamelli M. (Trebeschi et al., 2005). La documentazione cartografica per questo specifico territorio presenta una continuità che dai topografi napoleonici, passa a quella dell'Imperial Regio Istituto Geografico Militare Austriaco e, successivamente, a quello dell'Istituto Geografico Militare Italiano. Nella seconda metà dell'Ottocento, con la nascita dei i club alpini di Austria, Germania e Italia, inizia anche una produzione di ricerche scientifiche sulle Alpi.

In Italia l'organismo nazionale che dal 1895 cura la raccolta e la gestione dei dati sui ghiacciai è Comitato Glaciologico Italiano (CGI) (Smiraglia et al., 2021). Ne è testimonianza la pubblicazione del 1927 ad opera Porro e Labus del primo «Atlante dei Ghiacciai Italiani» (Firenze, IGM). In occasione dell'Anno Geofisico Internazionale (1957-58), il CGI ha realizzato, in collaborazione con il CNR, un secondo catasto dei ghiacciai italiani in 4 volumi (CGI-CNR 1959; 1962). Progetti e aggiornamenti effettuati da tali istituti hanno infine prodotto due importanti risultati che hanno riguardato la misurazione dei ghiacciai italiani effettuata con il programma CNR-NEXTD (2006/2007) (http://repo.igg.cnr.it/ghiacciaiCGI/ghiacciai_new.) e, la realizzazione del Nuovo Catasto dei Ghiacciai Italiani (NCGI) a cura di G. Smiraglia e G. Diolaiuti (2015)(htpp://sites.unimi.it/glaciol/index-php/it/catasto-dei-ghiacciai-italiani). Ne risulta un accurato aggiornamento delle misurazioni glaciologiche e dell'individuazione univoca degli apparati con possibilità di effettuare comparazioni con gli altri catasti.

La acclarata riduzione di tali sistemi è, quindi, un dato certo e le principali cause sono state l'aumento della temperatura media globale e la riduzione delle precipitazioni invernali.

Attualmente il Gruppo montuoso Adamello-Presanella, così definito per l'unitarietà geologica delle rocce ignee che lo caratterizzano, presenta 92 corpi glaciali (tavola1) ed è costituito dai due massicci: quello dell'Adamello e quello della Presanella, separati dalla Val di Genova nelle Alpi Retiche meridionali al confine tra la Regione Lombardia e la Provincia Autonoma di Trento. Nella tabella, che aggrega quanto pubblicato a scala regionale/provinciale nel Nuovo Catasto Glaciologico, vengono presentate le valutazioni rispetto alle catene montuose di pertinenza dei due enti territoriali.

Nel massiccio dell'Adamello si situa l'omonimo ghiacciaio, il più vasto delle Alpi italiane. È un apparato glaciale unitario classificato come un ghiacciaio di «altopiano con lingue radiali» dal quale si dipartono colate vallive che occupano la testata di alcune valli della Lombardia (Salarno e Adamè) e del Trentino (Val Genova). Il Gruppo montuoso dell'Adamello presenta, in territorio lombardo, 34 corpi glaciali. In tale area si trova la vedretta del Mandrone che presenta un elevato tasso di riduzione con un forte assottigliamento, criticità che, nell'agosto 2020, si sono manifestate con un collasso glaciale che ha creato, a una quota di circa 2600 m slm, un'imponente dolina circolare. Si stima che il volume del ghiaccio collassato sia di circa 100-120.000 m^3 (Centro nivometeorologico Arpa Lombardia). La situazione non è destinata a migliorare nonostante, in alcune aree del complesso glaciologico, si effettuino le coperture con teli geotessili e le pratiche di innevamento programmato a rinforzo della copertura nevosa.

Tavola 1. Distribuzione, area e variazioni dei ghiacciai lombardi e trentini suddivisi per gruppi montuosi, rispetto al Nuovo Catasto (NC), e al Catasto dei Ghiacciai Italiani (CGI),2015

Gruppo montuoso	N.ghiacciai (NC)	N.ghiacciai (CGI)	Area NC (km2)	Area CGI (km2)	Variazione numero	Variazione area (km2)	Variazione area (%)
Trentino							
Ortles-Cevedale	24	21	9.22	15.21	3	-5.99	-39%
Adamello-Presanella	58	46	19.05	25.44	12	-6.39	-25%
Brenta	20	16	0.86	3.23	4	-2.37	-73%
Marmolada	7	6	1.54	1.80	1	-0.26	-14%
Pale Di San Martino	6	2	0.29	0.79	4	-0.50	-63%
Tot.Trentino	115	91	30.96	46.47	24	-15.51	-33%
Lombardia							
Tambo’-Stella	12	13	1.83	3.85	-1	-2.02	-52%
Castello-Disgrazia	44	37	8.19	9.37	7	-1.18	-13%
Bernina-Scalino	21	14	21.27	26.98	7	-5.71	-21%
Livigno-Piazzi	26	26	4.23	7.08	0	-2.5	-40%
Ortles-Cevedale	51	42	28.58	42.44	9	-13.86	-33%
Adamello	34	34	21.62	23.29	2	-1.67	-7%
Orobie	42	21	1.95	1.85	22	0.10	5%
Tot.Lombardia	230	185	87.67	114.86	45	-27.19	-24%

Fonte: aggregazione delle tabelle del NCGI

3.1 Confini e governance

Il Gruppo montuoso con i suoi ghiacciai, come già accennato, si situa al confine tra la Regione Lombardia e la Provincia autonoma di Trento e subisce ulteriori ripartizioni. I massicci, infatti, ricadono sui territori limitrofi nel Parco Regionale Adamello (Lombardia) e nel Parco Naturale Adamello-Brenta (Trentino), e sono inseriti nella struttura ecologica di Rete Natura 2000 con i siti di interesse comunitario Sic IT 3120174 (Vermiglio-Folgaria); Sic Adamello IT3120175, la zona di protezione speciale ZPS Adamello-Presanella IT 3120158 e ZPS IT2070401 Parco Naturale Adamello.

Per l’area lombarda, quindi, il complesso del ghiacciaio dell’Adamello ricade nel territorio del Parco Naturale e in particolare la sua tutela è regolamentata anche alla scala europea.

Il Parco trentino fu istituito con la legge n. 18 del 1988. Successivamente il Ministero dell’Ambiente e della tutela del territorio e del mare ha, nelle more del recepimento della direttiva 92/43/CEE, approvato il V elenco aggiornato dei Siti di Importanza Comunitaria per la regione biogeografica alpina in Italia (D.L. 7 marzo 2012). Ne è risultata una ridefinizione dei perimetri di tali siti nella Provincia Autonoma di Trento con l'inserimento del versante Nord del Presena. Lo sviluppo delle Aree Protette si è stratificato su situazioni già in essere, come quelle delle concessioni d'uso del demanio sciabile.

In particolare, parte del complesso della Conca di Presena, particolarmente adatta alle pratiche sciistiche, è un grande circo glaciale aperto verso Nord Est delimitato dalle cime Monticello, Monticello superiore, Punta di Castellaccio, Corno di Lagoscuro, Cima di Presena, Cima Busazza. Inizialmente l'intero bacino era invaso dal ghiaccio ora rimangono le Vedrette di Presena occidentale, orientale e della Busazza occidentale ed orientale.

La peculiarità del Presena occidentale è la scarsa presenza di crepacci che ha favorito l'antropizzazione del ghiacciaio con il conseguente sviluppo dell'attività sciistica anche nei mesi estivi. (Carton, Tomasoni, Seppi, 2018). La storia dello sviluppo del settore turistico in tutta l'area è di lunga tradizione e fonte di strette relazioni tra l'alta Valle Camonica e la Valle di Sole. Separate dal Passo del Tonale, luogo di partenza degli impianti di risalita per il ghiacciaio, l'area è il fulcro dell'attrazione per le attività sciistiche ed escursionistiche. A tutt'oggi nel comprensorio Ponte di Legno/Tonale si contano 28 impianti che collegano tutte le piste tra Passo Tonale e Temù (uno tra questi impianti è quello che porta al Presena, cabinovia ammodernata a seguito del Programma di riordino del Ghiacciaio Presena, 2013).

La necessità di mantenere e tutelare il valore aggiunto che il settore turistico apporta all'economia per questi territori porta con sé il legame tra antropizzazione, ghiacciaio e buone pratiche che si sono implementate per la mitigazione dell'ablazione del ghiacciaio stesso. L'apparato glaciale del Presena Occidentale è oggetto di un programma di protezione attiva. Tale area non evolve più secondo le naturali condizioni ambientali, poiché dal 2008 si sta provvedendo alla copertura di parte del ghiacciaio mediante teli geotessili per limitare la fusione del ghiaccio e prolungarne la vita.

All'opzione della dismissione degli impianti sulla porzione interessata dalle attività sciistiche, si è preferita quella della adozione delle *best practices* nella gestione del ghiacciaio. Alla fine degli anni Ottanta queste hanno riguardato la sospensione delle attività sciistiche da giugno a settembre, per la tutela del ghiacciaio ma anche per l'impossibilità di praticare lo sci per gli effetti del cambiamento climatico, mentre a partire dal 2008 le amministrazioni si sono impegnate in azioni di protezione attiva con il Programma sperimentale «Mitigazione dell'ablazione e dell'impatto delle attività sciistiche del ghiacciaio Presena». Tale programma coinvolge la Provincia Autonoma di Trento tramite Meteotrentino, i comitati glaciologici nazionale e trentino, l'Università di Milano il Comune di Vermiglio e, con apposita convenzione con la Provincia, la società impiantistica Carosello Tonale SpA.

In estrema sintesi il Programma Sperimentale, prevede: la produzione di neve artificiale nella stagione fredda per incrementare quella derivante dagli apporti naturali; la riduzione dell'ablazione estiva con l'uso di teli geotessili posati in giugno e rimossi in settembre; la rilevazione dei principali parametri meteorologici, nivologici, glaciologici e idrometrici per quantificare il bilancio di massa e quello energetico del ghiacciaio, sia nelle zone protette dai teli che in quelle non protette; l'adozione di buone pratiche nella gestione dell'area sciistica (tecniche di battitura e spostamento della neve sulle piste, limitazioni all'apertura della pista in funzione della temperatura dell'aria e dello spessore del manto nevoso, ecc.).

Gli studi effettuati per la valutazione dei risultati di tale attività di mitigazione mostrano buoni risultati, in particolare hanno quantificato una riduzione complessiva dell'ablazione del 52% (periodo dello studio in esame si riferisce agli anni 2009, 2010, 2011) (Senese, Smiraglia, Diolaiuti, 2014). In aggiunta a tale Programma, sempre in ottica migliorativa, la Giunta provinciale della Provincia Autonoma di Trento, con la delibera n.405 del 8 marzo 2013, ha disciplinato gli interventi di manutenzione e razionalizzazione degli impianti e delle strutture esistenti, legati alla pratica dello sci sul ghiacciaio Presena. Ad orientare tali scelte vi è la considerazione dell'importanza economica per i comuni limitrofi al ghiacciaio delle attività derivanti dallo sci alpino (Valutazione Ambientale, 2012). La visione della sostenibilità risulta essere quella che promuove una montagna che sfrutti le proprie risorse rinnovabili, in accordo con quanto affermato dal comprensorio sciistico Ponte di Legno/Tonale, ovvero che «l'acqua dei torrenti (per le centraline idroelettriche) e la legna dei boschi (per la centrale del teleriscaldamento), produce una ricchezza che alimenta il motore del suo sviluppo» (https://www.pontedilegnoterme.it/press/), rispondendo pienamente alla lettura dell'Economia Ecologica.

Di contro guardando ai rapporti dell'osservatorio italiano del turismo montano e le ricerche sullo stesso tema, si ravvisa una certa cautela rispetto alla possibile ulteriore crescita del settore. Le osservazioni contenute in tali studi portano a riflettere sulle criticità e sugli effettivi benefici per le popolazioni locali dell'attuale modello di sviluppo turistico. In particolare, introducono un dubbio «sull'effettiva razionalità di tali investimenti, in relazione alle reali prospettive di mercato dell'economia sciistica, all'evoluzione dei redditi locali, al ritorno finanziario, alle conseguenze ambientali e al riscaldamento del clima» (CAI, 2020, p. 2).

L'uso del capitale naturale e la sua valorizzazione continuano a sollevare dibattiti sia nella narrazione che nell'interpretazione della sostenibilità poiché coniugare ambiente, economia e società potrebbe voler dire anche promuovere azioni che possano dare un futuro alle attività commerciali sui ghiacciai. La lettura utilitaristica più che quella solidale è la principale conclusione che emerge dall'analisi delle azioni svolte dagli attori locali.

Le riflessioni che emergono, quindi, mostrano come il concetto di utilità sul quale si basa il pensiero ecologia/economia, e dunque l'interpretazione funzionalista, non permetta di riconoscere alla Natura un valore intrinseco e dunque non le riconosce una posizione morale declassando la funzione della cura a pratiche cautelative.

4 Osservazioni conclusive

Nel trattare i punti oggetto di analisi si è ricostruito il ruolo che la biodiversità ha avuto nell'indirizzare il paradigma della sostenibilità verso l'unità di analisi ecosistemica. Cambia l'unità regionale, l'ecoregione, e si impone un maggiore coinvolgimento degli attori locali per la governance multilivello applicata alla gestione degli ambienti naturali che svolgono servizi sia per la scala locale che per quella globale con molteplici relazioni proprie dei sistemi complessi. L'approccio dell'Economia Ecologica inserisce infine nel modello le esternalità negative, quali la riduzione della biodiversità, come fallimenti del mercato. Attraverso il tema dei servizi offerti dagli ecosistemi e quello della misurazione della riduzione della capacità di svolgerli l'economia ha trovato la strada per creare un mercato per tali grandezze. Tale dinamica è stata evidenziata nella connessione tra le scelte della Strategia Nazionale sulla Biodiversità con quanto proposto e implementato alla scala europea e mondiale. Osservare quanto accade in uno specifico ambito ha permesso di visionare come si è incarnata questa chiave interpretativa della sostenibilità. In particolare, le unità ecologiche su cui valutare i servizi ecosistemici sono, come nel presente caso, situate nelle aree protette nelle quali la dinamica della conservazione diviene attiva (Nigel e al, 2011), come nel caso della copertura con i teli geotessili sul ghiacciaio del Presena occidentale, offrendoci un nuovo paesaggio che interiorizza le azioni di mitigazione.

L'Economia Ecologica riesce, dunque, ad introdurre nei modelli il valore economico della Natura, e, nonostante le difficoltà della misurazione dei servizi ecosistemici, a introdurre delle spinte correttive influenzando le decisioni rispetto alle strategie ambientali come nel caso in esame. La protezione attiva nel contesto esaminato contrasta nel breve termine la tendenza alla fusione dell'evoluzione naturale dei ghiacciai (cfr.§3) e garantisce, paradossalmente, proprio ciò che la sostenibilità voleva evitare, ossia la possibilità di un *business as usual* per gli imprenditori locali che si giovano dall'uso diretto del ghiacciaio. In questo è possibile riscontrare una lettura della sostenibilità agganciata alla visione funzionale del valore economico. Da queste considerazioni cosa si può apprendere? La mia inclinazione ottimista mi porta a guardare alle dinamiche neoliberiste che si disvelano nei valori di mercato date all'ambiente come a dei tutori che costringono, in qualche modo, a tenere conto del capitale naturale. A queste dinamiche si legano i temi e le implementazioni basate, ad esempio, sui pagamenti dei servizi ecosostemici (Landell-Mills e Porras, 2002; Poli, 2020). I dibattiti portati dall'interpretazione della *Social Nature*, allora, divengono quanto mai fondamentali non certo per trovare soluzioni ma per arricchire e dare spessore ai problemi che la società si trova ad affrontare nella dinamica Uomo/Natura. È condivisibile, a tal proposito, l'idea che la tutela della natura in sé non abbia senso se non in relazione con la vita dell'uomo (Zanolin, 2021), ma le compensazioni e le remunerazioni dei sevizi ecosistemici non risolvono, come mostrato dalle progressive ablazioni dei ghiacciai, le dinamiche del cambiamento. L'attesa e sperata transizione molto più radicale viene continuamente riproposta ma è sempre lecito domandarsi quali squilibri produrrà. L'apporto della geografia nell'arricchire il ragionamento attraverso le letture sui significati sociali insiti nell'uso, nella produzione, e riproduzione, della Natura è un mezzo per riconoscere le relazioni di interdipendenza e sviluppare maggiormente le valutazioni economiche degli impatti ambientali, aumentandone la complessità in modo da considerare anche gli effetti indotti delle eventuali azioni di mitigazione attiva come, rispetto al caso esaminato, le possibili valutazioni sulla quantità di CO_2 emessa per la produzione, il posizionamento, la conservazione e l'eventuale dismissione delle coperture con teli geotessili in termini di relazione tra diverse scale di analisi. Lo sguardo sulle dinamiche con cui si arriva alle policy di sviluppo locale e gli orientamenti dei portatori di interessi ne permette la valutazione critica. In questo risiede l'importanza della possibilità di esprimere letture differenti per sollecitare delle relazioni sostenibili maggiormente libere da impostazioni funzionali dove il ruolo della cura viene ridotto all'utilità. La lettura dei risultati delle dinamiche delle vedrette esaminate, infatti, mostra quantitativamente la capacità di effettuare azioni di tutela dei ghiacciai ma nel contempo mostra l'aumento degli squilibri. Se il ghiacciaio del Mandrone collassa sotto gli effetti del cambiamento climatico, il ghiacciaio del Presena occidentale riguadagna quel poco di consistenza che permette un ulteriore sfruttamento economico. Queste scelte, come detto, non fermano il riscaldamento globale: gli ampliamenti del demanio sciabile così come gli interventi di manutenzione e razionalizzazione propongono una lettura della cura, della tutela, e della sostenibilità in ottica dei diretti interessati agli investimenti, dimostrando di essere ben lontani dall'idea di società sostenibile. Si spiegano così gli appelli e le dichiarazioni di molte associazioni scientifiche e culturali come quella promossa dalla Carta dell'Adamello, a firma della Rete delle Università sostenibili del Club alpino italiano e del Comitato glaciologico italiano per la difesa dei ghiacciai, e il recente l'appello degli scienziati italiani (Gobbi et.al 2021). Dunque, la chiave dell'interesse economico permette di verificare le posizioni di forza e le conflittualità che la riproduzione sociale della Natura sta generando nelle aree protette esaminate. Le chiavi interpretative della geografia critica aiutano, invece, nella decodificazione del senso del prendersi cura e denunciano la mercificazione che la natura umana attua attraverso il paradigma economico.

Bibliografia

Barrett G., Farina A. (2000), Integrating ecology and economics, *BioScience*, 50 (4), 311-312. DOI: 10.1641/0006-3568(2000)050[0311:IEAE]2.3.CO;2.

Blasi C., Boitani L., La Posta S., Manes F., Marchetti M. (eds) (2009), *Contributo alla Strategia Nazionale per la biodiversità*, MINISTERO DELL'AMBIENTE DELLA TUTELA DEL TERRITORIO E DEL MARE (MATTM), Roma, Palombi Editori.

Blasi C., Capotorti G., Smiraglia D., Guida D., Zavattero L., Mollo B., Frondoni R., Copiz R. (eds) (2010)*, Le Ecoregioni d'Italia*, MINISTERO DELL'AMBIENTE DELLA TUTELA DEL TERRITORIO E DEL MARE (MATTM), Roma, Progetto Artiser.

M.Bagliai, A. Pietta (2022), Regione sostenibile e bioregione, in *Regioni e Regionalizzazione*, (eds) Gavinelli D., Bolocan Goldstein M., Pearson, Milano-Torino.

Bonati S., Tononi M., Zanolin G. (2021), Social nature geographies. Le geografie e l'approccio sociale alla natura*, Rivista Geografica Italiana*, 2, 5-21. DOI 10.3280/rgioa2-2021oa12029.

Brown L. R. (2001), *Eco-Economy: Building an Economy for the Earth*, New York, W.W Norton & Co.

Capriolo A. (2020), *Il sostegno all'economia dato dal capitale naturale*, in «Giornata Mondiale Della Biodiversità», 22 Maggio 2020 Giornata Italiana Del Capitale Naturale, Convegno Scientifico, available at https://www.minambiente.it/pagina/giornata-mondiale-della-biodiversita-2020. [30 Marzo 2022].

Carton A., Tomasoni R., Seppi R. (2018), *Dal Passo del Tonale al Lago di Tovel, nel cuore dell'Adamello* – Brenta Geopark, available at http://www.europeangeoparks.org. [30 Marzo 2022].

CLUB ALPINO ITALIANO, COMMISSIONE CENTRALE TUTELA AMBIENTE E TERRITORIO MONTANO, (2020), *Cambiamenti Climatici, Neve, Industria dello Sci dello sci. Analisi Del Contesto, Prospettive e Proposte*, https://www.caiuget.it/tam/cambiamenti-climatici-neve-industria-dello-sci-analisi-del-contesto-prospettive-e-proposte.html. [30 Marzo 2022].

COMITATO PER IL CAPITALE NATURALE ITALIANO (CCN) (2020), Convegno Scientifico: Conoscere, Valutare, Conservare e Arricchire il Capitale Naturale In Italia. Available at https://www.minambiente.it/pagina/giornata-mondiale-della-biodiversita-2020. [30 Marzo 2022].

COMMISSIONE EUROPEA, COM (2011) 21 def., *Un'Europa efficiente nell'impiego delle risorse – Iniziativa faro nell'ambito della strategia Europa* 2020, Bruxelles.

COMMISSIONE EUROPEA COM (2012) 710 final 2012/0337, *Vivere bene entro i limiti del nostro pianeta*, Bruxelles.

COMMISSIONE EUROPEA, COM (2013) 249 final, *Infrastrutture verdi – Rafforzare il capitale naturale in Europa*, Bruxelles.

CONFERENCE OF THE PARTIES TO THE CONVENTION ON BIOLOGICAL DIVERSITY (2010), *Decision adopted by the conference of the parties to the Convention on Biological Diversity at its tenth meeting, Nagoya, Japan, Agenda item 3*.

COUNCIL OF THE EUROPEAN UNION (2019) 12795/19, *The 8th Environment Action Programme - Turning the Trends Together*, Brussels.

Costanza R. (1989), What is ecological economics?, *Ecological Economics*, 1, 1-7. DOI: 10.1016/0921-8009(89)90020-7.

Costanza R., D'arge R., De Groot R., Faber S., Grasso M., Hannon B., Limburg K., Naeem S., O'neill R., Paruelo J., Raskin R., Sutton P., Van Den Belt M. (1997), The value of the world's ecosystem services and natural capital, *Nature*, 387, 253-260.

Daly H. E. (1977), *Steady State Economics*, Whashington, Island Press, (trad.it. 1981: *Lo stato stazionario*, Firenze, Sansono).

Daly H. E. (1997), Georgescu- Roegen versus Solow/Stiglitz, *Ecological Economics*, 22(3), 261-266. DOI:10.1057/9780230523210_13.

EUROPEAN ENVIRONMENT AGENCY (2017), *Environmental indicator report 2017 In support to the monitoring of the Seventh Environment Action Programme*, Luxembourg: Publications Office of the European Union. DOI:10.2800/905696

Folke C., Carpenter S., Walker B., Scheffer M., Chapin T., Rockström J. (2010), Resilience thinking: integrating resilience, adaptability and transformability, *Ecology and Society*, 15 (4), 20. available at http://www.ecologyandsociety.org/vol15/iss4/art20/ [30 Marzo 2022]. DOI:10.5751/ES-03610-150420.

Giupponi C., Galassi S., Pettanella D. (2009), *Definizione del metodo per la classificazione e quantificazione dei servizi ecosistemici in Italia*, MINISTERO DELL'AMBIENTE DELLA TUTELA DEL TERRITORIO E DEL MARE (MATTM).

Gobbi M., Ambrosini R., Casarotto C., Diolaiuti G., Ficetola F., Leoncini V., Seppi R., Smiraglia C., Tampucci D., Valle B., Caccianiga M. (2021), Vanishing permanent glaciers: climate change is threatening a European Union habitat (Code 8340) and it's poorly biodiversity, *Biodiversity and Conservation*, 30, 2267-2276. DOI:10.1007/s10531-021-02185-9.

Gómez-Baggethun E., De Groot R., Lomas P. L., Montes C. (2010), The history of ecosystem services in economic theory and practice: From early notions to markets and payment schemes, *Ecological Economics*, 69(6), 1209-1218. DOI: 10.1016/j.ecolecon.2009.11.007.

INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (IPCC) (2019), Special Report on the Ocean and Cryosphere in a Changing Climate.

Landell-Mills N., Porras I.T. (2002), *Silver bullet or fools' gold? A global review of markets for forest environmental services and their impact on the poor*, International Institute for Environment and Development, London.

MILLENNIUM ECOSYSTEM ASSESSMENT (2005), *Ecosystems and human well-being: the assessment series (4 vol + Summary)*, Washington DC., Island Press.

MINISTERO DELL'AMBIENTE DELLA TUTELA DEL TERRITORIO E DEL MARE (MATTM) (2010), *La Strategia Nazionale per la Biodiversità*.

MINISTERO DELL'AMBIENTE DELLA TUTELA DEL TERRITORIO E DEL MARE (MATTM) (2012), *Strategia Nazionale per la Biodiversità, 1°Rapporto (2011-2012)*.

MINISTERO DELL'AMBIENTE DELLA TUTELA DEL TERRITORIO E DEL MARE (MATTM) (2014), *Strategia Nazionale per la Biodiversità, Secondo Rapporto 2013-2014*.

MINISTERO DELL'AMBIENTE DELLA TUTELA DEL TERRITORIO E DEL MARE (MATTM) (2016), *Strategia Nazionale per la Biodiversità,Terzo Rapporto 2015-2016.*

MINISTERO DELL'AMBIENTE DELLA TUTELA DEL TERRITORIO E DEL MARE (MATTM) (2017), *Comitato per il Capitale Naturale, Primo Rapporto sullo stato del Capitale Naturale in Italia*.

MINISTERO DELLA TRANSIZIONE ECOLOGICA (2021), *Comitato per il Capitale Naturale, Quarto Rapporto sullo stato del Capitale Naturale in Italia.*

MINISTERO DELL'ECONOMIA E DELLE FINANZE (2021), *Piano Nazionale di Ripresa e Resilienza (PNRR).*

Nelson J., Power M. (2018), Ecology, Sustainability, and Care: Developments in the Field, *Feminist Economics*, 24(3), 80-88. DOI:10.1080/13545701.2018.1473914

Nigel D., Higgins-Zogib L., Hockings M., Mackinnon K., Sandwith T., Stolton S. (2011), National Parks with benefits: how protecting the planet's biodiversity also provides ecosystem services, *Solutions*, 2 (6), 87-95.

O'hara S. (1999), Economics, Ecology, and Quality of Life: Who Evaluates?, *Feminist Economics*, 5(2), pp.83-89. DOI: 10.1080/135457099337969.

ORGANIZZAZIONE DELLE NAZIONI UNITE (2015), *Trasformare il nostro mondo: l'Agenda 2030 per lo Sviluppo Sostenibile, 4 ° riunione plenaria*.

Poli D. (2020*), I servizi ecosistemici nella pianificazione bioregionale*, Firenze University Press, Firenze.

PROVINCIA AUTONOMA DI TRENTO (2012), *Valutazione Ambientale Strategica del Programma degli interventi di manutenzione e razionalizzazione degli impianti e delle strutture esistenti, legati alla pratica dello sci, e degli interventi di valorizzazione ambientale e culturale, anche ai fini turistici. Rapporto Ambientale*.

Rockström J., Steffen W., Noone K., Persson A., Chapin F. S. III, Lambin E.F., Lenton T.M., Scheffer M., Folke C., Schellnhuber H.J., Nykvist B., De Wit C.A., Hughes T., Van Der Leeuw S., Rodhe H., Sörlin S., Snyder P.K., Costanza R., Svedin U., Falkenmark M., Karlberg L., Corell R.W., Fabry V.J., Hansen J., Walker B., Liverman D., Richardson K., Crutzen P., Foley J.A. (2009), A safe operating space for humanity, *Nature*, 461, 472–475. DOI: 10.1038/461472a.

Rossi U. (2008), Il post-sviluppo nel Sud globale: sfida possibile o nuova illusione?, in Boggio F., Dematteis G., Memoli M. (eds.), *Geografia dello Sviluppo*, UTET, Torino.

Senese A., Smiraglia C., Diolaiuti G., (2014), *Studi pilota per l'utilizzo dei teli geotessili nontessuti per la riduzione della fusione della neve e del ghiaccio. Risultati del Ghiacciaio del Presena Occidentale (Trento)*, Pàtron, Bologna.

Smiraglia C., Fugazza D., Diolaiuti G. (2021), Continua inarrestabile il regresso dei ghiacciai italiani e alpini. Le evidenze dei recenti catasti, *Il Bollettino*, CAI.

Smiraglia C. and Diolaiuti G. (eds) (2015), *Nuovo Catasto dei Ghiacciai Italiani (NCGI)*, available at htpp://sites.unimi.it/glaciol/index-php/it/catasto-dei-ghiacciai-italiani [30 Marzo 2022].

Steffen W., Crutzen P. J., Mcneill J. R. (2007), The Anthropocene: Are Humans Now Overwhelming the Great Forces of Nature?, *Ambio*, 36 (8),614-621. DOI: 10.1579/0044-7447(2007)36[614:taahno]2.0.co;2.

Trebeschi A., Guzzetti F., Vassena G., Vavassori F. (2005), Adamello: la variazione di quota della Vetta, in due secoli di misurazioni, *Atti Asita*, *http://atti.asita.it›* Asita2005 ›pdf [30 Marzo 2022]..

Waring M. (1988), *If Women Counted: A New Feminist Economics*, CA:Harper & Row, San Francisco.

Zanolin G. (2021), La natura e l'immaginario: le aree protette come costruzioni sociali, *Rivista Geografica Italiana*, 2, 85-101. DOI: 10.3280/rgioa2-2021oa12034.

SAPIENZA
UNIVERSITÀ EDITRICE

ANNALI DEL DIPARTIMENTO DI METODI E MODELLI
PER L'ECONOMIA IL TERRITORIO E LA FINANZA
2022 | pp. 107-118

E-ISSN 2611-6634 | ISSN 2385-0825
DOI: 10.13133/2611-6634/1410

Research article

Cinzia Di Palo*, **Pierluigi Fava**°

ANALISI STOCASTICA DI UNA EMISSIONE IN VALUTA

Abstract.
This study analyses a potential investment by a possible pension investor in government bonds issued in foreign currency under a risk-return perspective. In particular, the government bond under study is the thirty-year Italy Govt ISIN US465410CC03, denominated in US dollars. By means of stochastic processes, the trend in the exchange rate between the two currencies the dollar and the euro is simulated to assess its impact on the yield to maturity of the government bond considered. The analysis involves determining the probabilities of obtaining a negative yield in the different simulation scenarios. In addition, there are also determined the probabilities of obtaining returns to maturity lower than those obtainable with a similar investment instrument denominated in euro.

Keywords: exchange rates, government bond, stochastic simulation.

---

* Università degli studi di Cassino e del Lazio Meridionale
° (Pierluigi Fava) Sapienza Università di Roma

1 Introduzione

Il presente studio analizza sotto il profilo rischio-rendimento il potenziale investimento in titoli governativi emessi in valuta estera da parte di un Investitore Istituzionale di natura previdenziale. In particolare, l'analisi ha per oggetto il titolo trentennale Italy Govt ISIN US465410CC03 emesso il 27 Aprile 2021 denominato in dollari statunitensi. Si ipotizza la natura previdenziale dell'Investitore Istituzionale e si assume che questo, in ottica di cash flow matching, detenga il titolo sino a scadenza e che, quindi, le variazioni di prezzo del titolo, apprezzate nel suo arco di vita, non ne influenzino il rendimento.

Il lavoro si sviluppa nel modo seguente. Nella sezione 2 è descritto il titolo Italy Govt ISIN US465410CC03, denominato in dollari statunitensi, oggetto del nostro studio e sono descritti i rischi connessi all'impatto che le variazioni del rapporto di cambio hanno sul titolo stesso. Nella sezione 3, attraverso processi stocastici, è sviluppata la simulazione dell'andamento del rapporto di cambio tra dollaro statunitense ed euro al fine di valutare come questo influenzi il rendimento a scadenza del titolo in esame. Infine, nella sezione 4 sono sintetizzati i principali risultati dell'analisi sviluppata.

2 Analisi dell'Emissione governativa

Con nota del 27 Aprile 2021 il Dipartimento del Tesoro del Ministero dell'Economia e delle Finanze italiano annunciava l'emissione per 1500 milioni di dollari statunitensi (USD) del titolo trentennale Italy Govt ISIN US465410CC03 (Italy Tf 3.875% Mg51 Usd), scadenza 6 maggio 2051 e tasso cedolare del 3,875% corrisposto con periodicità semestrale. Al prezzo di collocamento di 98,89 USD, il titolo corrispondeva un rendimento lordo all'emissione del 3,938%. Il titolo è classificato come debito senior emesso dalla Repubblica Italiana.

L'asta ha registrato domanda per 6,2 miliardi di dollari statunitensi con collocamento sindacale.

Se si confronta il rendimento a scadenza del titolo oggetto di analisi con un analogo strumento emesso in valuta domestica, per esempio il titolo trentennale ISIN IT0005425233 (Btp Tf 1,7% St51 Eur), scadenza 1 settembre 2051 e tasso cedolare dell'1,7%, anch'esso corrisposto con periodicità semestrale e classificato con medesima seniority, si osserva alla data di emissione del titolo in valuta, a parità di rischio di emittente, un differenziale in termini di rendimento annuale a scadenza di circa 200 bps (fonte Bloomberg) dovuto ai differenti tassi espressi dalla curva del Treasury USA[1].

In aggiunta, rispetto al Treasury trentennale classificato con rating AAA, il titolo Italy Govt ISIN US465410CC03 emesso in dollari statunitensi registra uno spread all'emissione di circa 160 bps (fonte Bloomberg). Tale spread riflette il rischio di credito della Repubblica Italiana rispetto agli Stati Uniti.

Tali differenziali suggeriscono di approfondire il profilo rischio-rendimento del titolo in esame. In generale, un investimento in un titolo denominato in valuta estera determina per il sottoscrittore l'esposizione al rischio di cambio. La posizione lunga su un titolo emesso in valuta estera è soggetta al rischio di un rafforzamento della valuta

---

[1] Sebbene il titolo sia emesso dalla Repubblica Italiana, una emissione in valuta estera deve considerare la curva dei rendimenti espressa dal Paese di riferimento della denominazione.

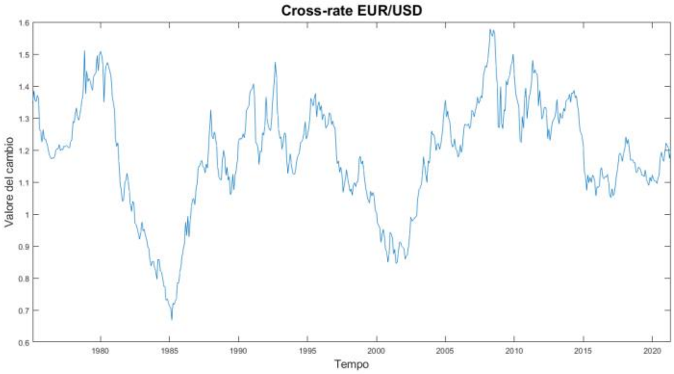
domestica, che si tradurrebbe in una riduzione del tasso interno di rendimento per via della riduzione degli importi associati al cash flow.

Pertanto, al fine di valutare il margine di profittabilità dell'investimento in funzione del rischio di cambio, analizziamo il Cross-rate Euro/Dollaro (EUR/USD), che coinvolge le valute delle due più sviluppate economie globali e risulta essere, in termini di volumi di negoziazione, la prima asset class su scala mondiale.

2.1 Analisi del Cross-rate Euro/Dollaro

La serie storica del rapporto di cambio EUR/USD dal 1975 ad oggi mostra un andamento ciclico, vedi Figura 1. In particolare, si registra un minimo di 0.66 a febbraio 1985 ed un massimo di 1.58 a marzo 2008.

Figura 1. Rapporto di cambio EUR/USD dal 1975 al 2021, con frequenza di rilevamento mensile (*valutazione a fine mese*)



Fonte: Elaborazione propria su dati Bloomberg

Secondo la teoria della Parità del Potere d'Acquisto (PPA) (Terborgh, 1926) è possibile stabilire se il livello corrente del Cross-rate sia allineato o meno al valore di equilibrio. La PPA definisce una condizione di assenza di arbitraggio affermando che due panieri di beni identici, venduti in paesi diversi, devono avere lo stesso prezzo una volta espressi in valuta comune.

Come dall'analisi del provider Bloomberg, con riferimento ai valori osservati a maggio 2021, si osserva che il Cross-rate EUR/USD appare sostanzialmente in equilibrio in quanto sul mercato si registra un rapporto di cambio EUR/USD pari a 1.22 (il valore teorico derivante dalla teoria della PPA è 1.20).

In virtù delle osservazioni di cui sopra, ritenendo "mature" entrambe le economie (europea e statunitense), si presuppone che eventuali oscillazioni anomale, sia in ampiezza che in velocità, del rapporto di cambio EUR/USD vengano considerate dalle rispettive Banche Centrali al fine di evitare il diffondersi di shock sull'economia reale.

3 Analisi quantitativa del rischio di cambio associato all'investimento

Nel prosieguo del lavoro viene considerata la serie storica del rapporto di cambio USD/EUR, in modo che un apprezzamento del dollaro rispetto all'euro si traduca in un rendimento positivo sull'investimento nel titolo Italy Govt ISIN US465410CC03 in dollari statunitensi.

In questa sezione, ai fini della valutazione del rischio di cambio associato all'investimento, si presenta prima l'analisi statistico/econometrica sulla serie storica del rapporto di cambio USD/EUR e quindi si descrivono brevemente i processi stocastici implementati e le relative calibrazioni per la simulazione del trend futuro del rapporto di cambio USD/EUR. Infine, si illustra la metodologia adottata per la valutazione del rischio di cambio sul titolo in esame.

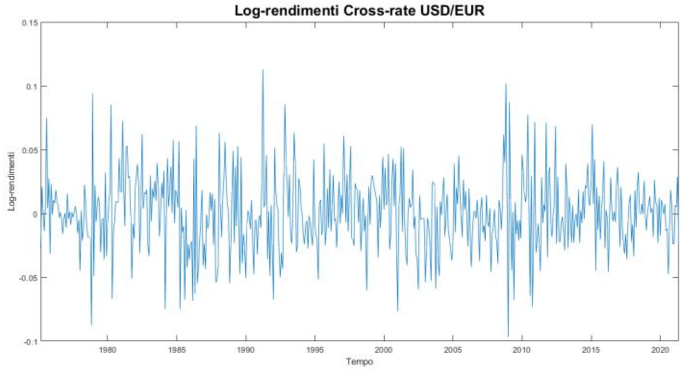
3.1 Analisi statistico/econometrica preliminare

Dal provider finanziario Bloomberg è stata ottenuta la serie storica del rapporto di cambio USD/EUR, con frequenza mensile, dal 31/01/1975 al 30/04/2021. La serie storica si compone di 556 osservazioni. Sono stati, quindi, calcolati i log-rendimenti del rapporto di cambio USD/EUR, con la formula

$$r_i = \ln\left(\frac{x_i}{x_{i-1}}\right) \qquad \text{con } i = 1,2,\dots 555 \qquad (1)$$

in cui x_i indica il rapporto di cambio USD/EUR al tempo i e x_{i-1} il valore dello stesso al mese precedente, vedi Figura 2.

Figura 2. Serie storica dei rendimenti logaritmici del rapporto di cambio USD/EUR dal 31 gennaio 1975 al 30 aprile 2021

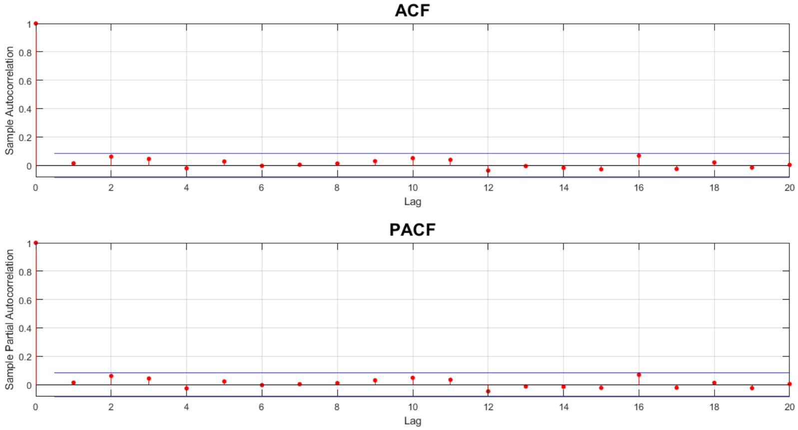


Fonte: Elaborazione propria su dati Bloomberg

La serie storica in esame presenta parametri di asimmetria e curtosi rispettivamente pari a 0.20 e 3.9, valori leggermente superiori rispettivamente a 0 e 3 caratteristici di

una distribuzione normale. Il risultato del test Jarque-Brera (1987) conferma, con p-value al 5%, il rifiuto della distribuzione normale per i rendimenti.
L’analisi dei residui, effettuata tramite rappresentazione grafica della funzione di autocorrelazione (ACF) e della funzione di autocorrelazione parziale (PACF), non mostra segni di autocorrelazione, vedi Figura 3. Il test econometrico di Ljung-Box (1978) con p-value al 5% indica di non poter rifiutare l’ipotesi nulla di non autocorrelazione dei residui.

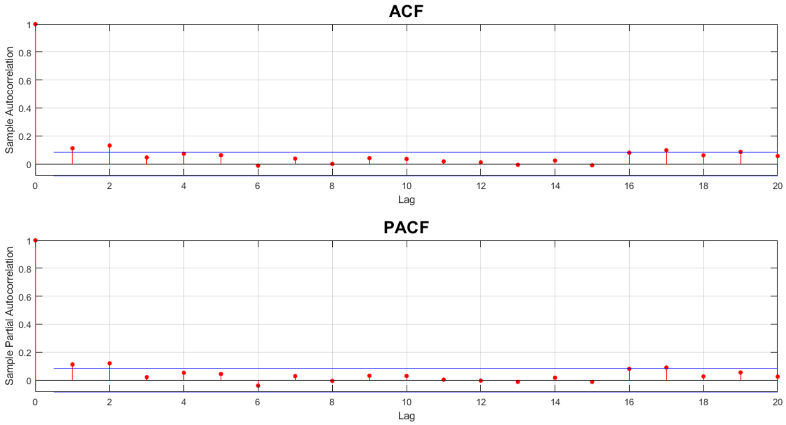
Figura 3. Rappresentazione della ACF e della PACF dei residui



Fonte: Elaborazione propria

Dall’analisi dei residui al quadrato, vedi Figura 4, si osserva la presenza di autocorrelazione in prossimità del primo e secondo lag, in cui si evidenzia una rottura delle bande di supporto, la quale suggerisce la presenza di volatility cluster.

Figura 4. Rappresentazione della ACF e della PACF del quadrato dei residui



Fonte: Elaborazione propria

Il test econometrico di Engle (1982) sui residui con p-value al 5% suggerisce la presenza di omoschedasticità nella volatilità.

3.2 Presentazione dei processi stocastici e calibrazioni

Questa sottosezione fornisce una breve descrizione dei processi stocastici utilizzati per simulare l'andamento del rapporto di cambio USD/EUR, al fine di valutarne il possibile impatto sul rendimento del titolo obbligazionario Italy Govt ISIN US465410CC03.

Per la descrizione dettagliata dei processi stocastici e delle relative funzioni di verosimiglianza utilizzate per la stima dei parametri si rimanda a Brigo et al. (2017).

3.2.1 Moto Browniano Geometrico

Il moto Browniano geometrico è il processo utilizzato per descrivere il comportamento casuale nel tempo del prezzo di un'azione (Hull, 2009). Indicato con S_t il prezzo dell'azione al tempo t, l'equazione differenziale stocastica che descrive il processo risulta essere:

$$dS_t = \mu S_t \mathrm{dt} + \sigma S_t dW_t$$

dove i parametri μ (tasso di deriva) e σ (tasso di varianza) rappresentano, rispettivamente, il tasso di rendimento atteso e la volatilità del prezzo dell'azione, e dW_t è un moto Browniano standard (anche noto come processo di Wiener) caratterizzato da incrementi mutuamente indipendenti e distribuiti normalmente con media zero e deviazione standard pari alla radice quadrata dell'incremento del tempo.

L'equazione continua del processo Browniano geometrico considerata tra due istanti discreti, t_i e t_{i+1}, ha per soluzione:

$$S_{t_{i+1}} = S_{t_i} e^{\left(\mu - \frac{\sigma^2}{2}\right)(t_{i+1} - t_i) + \sigma Z_{i+1}\sqrt{t_{i+1} - t_i}}$$

dove $Z_1,\ Z_2, \dots Z_n$ sono estrazioni casuali da una distribuzione normale standard.

Nella nostra simulazione, la variabile S_t denota il rapporto di cambio USD/EUR al tempo t. L'intervallo di tempo considerato è $[t_0,\ t_n]$, con t_0 e t_n fissati rispettivamente al 30 aprile 2021 e 31 maggio 2051. Poiché si considerano rilevazioni mensili del rapporto di cambio, gli istanti di tempo discreto t_i sono scelti in modo tale che $\Delta t = t_{i+1} - t_i = 1/12$ di anno per ogni i.

Per la stima dei parametri del processo, è stato utilizzato il metodo di massima verosimiglianza. Per questo processo, gli incrementi dei log-rendimenti sono variabili casuali normali, indipendenti e identicamente distribuite, ciascuna con funzione densità con media m e varianza v. Gli stimatori dei parametri del processo del moto Browniano geometrico sono calcolati come segue:

$$m = \left[\hat{\mu} - \tfrac{1}{2}\widehat{\sigma^2}\right]\Delta t \qquad\qquad v = \widehat{\sigma^2}\Delta t$$

Pertanto, questi sono ottenuti dalle stime dei parametri m e v. Per il processo Browniano geometrico, il metodo di massima verosimiglianza fornisce formule ben note in forma chiusa per media e varianza del campione dei log-rendimenti (vedi Brigo et al. 2007).

La funzione di log-verosimiglianza, denotata come $L^*(\mu, \sigma)$, è definita come:

$$L^*(\mu, \sigma) = n\, log\left(\frac{1}{\sigma\sqrt{2\pi dt}}\right) + \sum_{i=1}^{n} \frac{-\left(r_i - (\mu - \frac{\sigma^2}{2})dt\right)^2}{2\sigma^2 dt}$$

dove n è il numero dei log-rendimenti mensili osservati, con $n = 555$, r_i indica l'i-esimo log-rendimento mensile osservato, vedi formula (1), e dt è posto uguale a $\frac{1}{12}$ di anno.

I valori stimati per i parametri del processo, espressi su base annua, sono rispettivamente:

- $\hat{\mu} = 0.0077$
- $\hat{\sigma} = 0.1021$

3.2.2 Processo Variance-Gamma

Il secondo processo stocastico considerato è il Variance-Gamma (VG), presentato da Madan e Seneta (1990), è utilizzato per modellare "fat tails" nella distribuzione dei rendimenti. Si definiscono "fat tails" i fenomeni in cui grandi variazioni numeriche estreme sono più frequenti di quanto mostrato da una ipotetica distribuzione normale.

Sia S_t la variabile che denota il rapporto di cambio USD/EUR al tempo t. Il processo VG segue la seguente dinamica:

$$dlogS_t = \mu dt + \theta dg_t + \sigma dW(g_t)$$

dove μ, θ e σ sono costanti reali, con $\sigma \geq 0$. Rispetto al moto Browniano geometrico si evidenzia la presenza del termine g_t che è un processo stocastico crescente e positivo, con incrementi stazionari e indipendenti. Il processo VG assume che g_t segua un processo Gamma con parametro ν indipendente dal moto Browniano standard $\{W_t\}_{t \geq 0}$.

Denotiamo con $\pi = \{\mu, \theta, \sigma, \nu\}$ l'insieme dei parametri della seguente funzione densità di probabilità variance-gamma:

$$f(X_{\Delta t})(x) = \int_0^\infty f_N(x; \hat{\theta} g; \hat{\sigma}^2 g) f_\Gamma\left(g; \frac{\Delta t}{\nu}, \nu\right) dg$$

dove $\Delta t = t_{i+1} - t_i$ e $X_{\Delta t} = \log\left(\frac{S(t+\Delta t)}{S(t)}\right)$.

La stima di tali parametri a partire delle serie storica dei log-rendimento del rapporto di cambio USD/EUR è effettuata con il metodo di massima verosimiglianza. I valori di inizializzazione della calibrazione sono ottenuti tramite il metodo della funzione generatrice dei momenti:

$$M_X(z) = e^{\hat{\mu} z}(1 - \hat{\theta} \nu z - \frac{1}{2}\nu \hat{\sigma}^2 z^2)^{-\frac{\Delta t}{\nu}}$$

Si rimanda a Brigo et al. (2017) per la descrizione della procedura completa di stima.

I valori stimati per i parametri del processo, espressi su base annua, sono rispettivamente:

- $\hat{\mu} = -0.006$
- $\hat{\sigma} = 0.0295$
- $\hat{\nu} = 0.32$
- $\hat{\theta} = 0.0062$

3.2.3 Modello di Vasicek

Il terzo processo considerato è quello di Vasicek (1977), uno dei primi modelli stocastici per il tasso di interesse a breve termine. Questo processo è caratterizzato dalla proprietà di mean reverting, proprietà che appare nell'osservazione della serie storica del rapporto di cambio USD/EUR.

Il modello è descritto dalla seguente equazione differenziale stocastica:

$$dx_t = \alpha(\theta - x_t)dt + \sigma dW_t \tag{2}$$

in cui x_t indica il tasso di cambio al tempo t, i parametri α (velocità di ritorno alla media), θ (livello tendenziale) e σ (volatilità) sono positivi e dW_t descrive un moto Browniano standard.

La versione a tempo discreto della soluzione dell'equazione (2) è:

$$x_{t_i} = \theta\left(1 - e^{-\alpha\Delta t}\right) + e^{-\alpha\Delta t}x_{t_{i-1}} + \sigma\sqrt{\frac{1 - e^{-2\alpha\Delta t}}{2\alpha}}\epsilon_{t_i}$$

dove ϵ_t è la realizzazione al tempo t di una variabile aleatoria normale standard ed è posto $\Delta t = t_i - t_{i-1}$.

La calibrazione del processo avviene tramite regressione lineare a partire dalle differenze prime delle osservazioni mensili dei tassi di cambio.

I valori stimati per i parametri del processo, espressi su base annua, sono rispettivamente:

- $\hat{\alpha} = 0.1947$
- $\hat{\theta} = 0.8745$
- $\hat{\sigma} = 0.09$

3.2.4 Modello di Vasicek di tipo esponenziale

Il quarto processo considerato è quello di Vasicek di tipo esponenziale, che è utilizzato, oltre che per la proprietà di mean reverting, anche per tener conto della positività dei tassi di interesse e catturare eventuali fat tails nella loro distribuzione.

Sia $y(t)$ descritta da un processo di Vasicek. Posto $x(t) = exp\left(y(t)\right)$, in modo da garantire la positività della variabile, il modello di Vasicek esponenziale è descritto dalla seguente equazione differenziale stocastica:

$$dx_t = \alpha x_t(m - log\ (x_t))dt + \sigma x_t dW_t \tag{3}$$

in cui α e σ sono costanti positive, $m = \theta + \frac{\sigma^2}{2\alpha}$ e dW_t è un moto Browniano standard.

La soluzione esplicita della equazione (3), considerata tra due qualsiasi istanti s e t, con $0 < s < t$, è data da:

$$x_t = \exp\left(\theta\left(1 - e^{-\alpha(t-s)}\right) + \log(x_s)e^{-\alpha(t-s)} + \sigma e^{-\alpha t}\int_s^t e^{\alpha u} dW_u\right)$$

con dW_t moto Browniano standard.

La calibrazione del processo x_t avviene mediante calibrazione del suo processo logaritmico tramite regressione lineare a partire dalle differenze prime delle osservazioni mensili dei tassi di cambio USD/EUR.

I valori stimati per i parametri del processo, espressi su base annua, sono rispettivamente:

- $\hat{\alpha} = 0.219$
- $\hat{\theta} = -0.1473$
- $\hat{\sigma} = 0.1017$

3.3 Metodologia per la valutazione del rischio di cambio

La metodologia adottata prevede di calcolare i rendimenti a scadenza dell'investimento del titolo Italy Govt ISIN US465410CC03, denominato in dollari, mediante simulazioni stocastiche. Sono quindi determinate le corrispondenti probabilità di shortfall, definite come:

a) le probabilità che il titolo ottenga un rendimento a scadenza negativo;

b) le probabilità che il titolo ottenga un rendimento a scadenza inferiore al rendimento a scadenza del titolo ISIN IT0005425233, denominato in euro, preso a confronto.

In aggiunta, per ognuno dei processi simulati, sono forniti i percentili dei rendimenti a scadenza del titolo.

Nello specifico, il titolo Italy Govt ISIN US465410CC03 prevede il pagamento di cedole semestrali posticipate il 6 maggio e il 6 novembre di ogni anno a partire dal 6 novembre 2021 e fino al 6 maggio 2051. Pertanto, il numero totale dei flussi è pari a 61 (60 flussi sono le cedole in incasso per l'investitore, compreso il rimborso finale, ed un flusso è dato dal pagamento per l'investitore alla data iniziale di acquisto del titolo).

Nella nostra simulazione, ad ogni data prevista d'incasso, ovvero, ad ogni data d'incasso della cedola ed alla data di rimborso a scadenza del titolo, il cash flow in dollari del titolo è valutato in base al valore del tasso di cambio USD/EUR per ogni scenario simulato. Il numero degli scenari della simulazione è pari a 10.000.

Le valutazioni dei possibili cash flow sono espresse in euro sulla base del valore simulato del tasso di cambio USD/EUR.

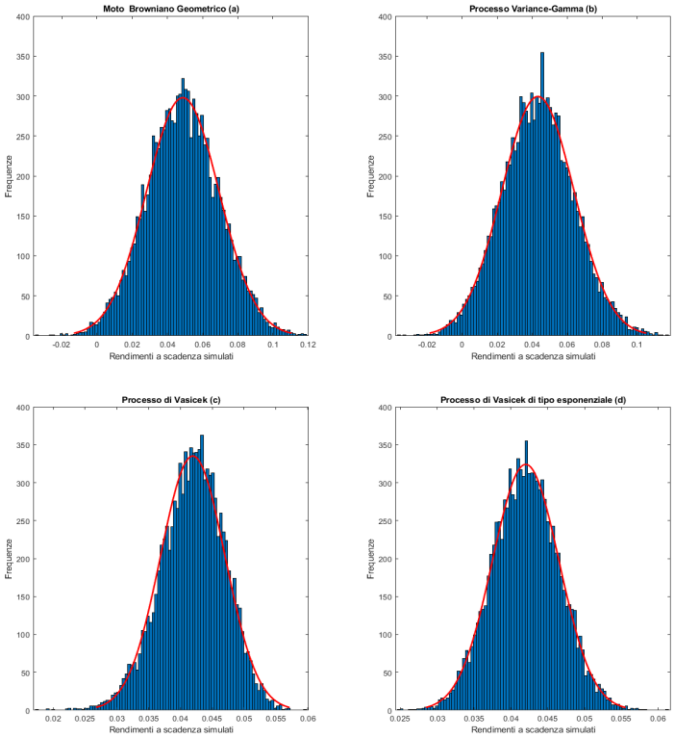
I passi delle simulazioni sono mensili.

Le quotazioni di input del modello si riferiscono al 14 maggio 2021 data in cui il titolo Italy Govt ISIN US465410CC03 quotava 98,90 USD (prezzo dirty ask) con un rendimento a scadenza di circa 3,94% (fonte Bloomberg). Alla stessa data, il corrispondente titolo ISIN IT0005425233ISIN, denominato in valuta domestica, presentava un rendimento a scadenza dell'1,95% (fonte Bloomberg).

4 Risultati delle simulazioni e principali conclusioni

Nella Figura 5 sono riportate le distribuzioni dei rendimenti a scadenza ottenute mediante la simulazione con le 10.000 traiettorie del rapporto di cambio USD/EUR generate per ciascuno dei processi stocastici considerati.

Figura 5. Distribuzione dei rendimenti a scadenza simulati, per processo generatore dei dati



Fonte: Elaborazione propria

Le probabilità di shortfall sono riportate nella Tabella 1. Si osserva che la probabilità di registrare un rendimento a scadenza negativo risulta massima nella simulazione con il processo VG (pari a 1,60%), si attesta al di sotto dell'1% nella simulazione con il moto Browniano geometrico, ed è nulla nelle altre due simulazioni. Analogamente, le probabilità di conseguire performance inferiori a quelle del corrispondente titolo in valuta domestica risulta massima nella simulazione con il processo VG (pari a 13,09%), si attesta a circa l'8% nella simulazione con il moto Browniano standard, risulta pari allo 0,03% nella simulazione con il processo di Vasicek ed è nulla nelle simulazioni con il processo di Vasicek di tipo esponenziale.

Tabella 1. Probabilità di shortfall, per processo generatore dei dati e per tipo di rendimento a scadenza

Processo generatore dei dati	Shortfall Probability	
	Rendimento a scadenza minore di zero	Rendimento a scadenza del titolo in dollari minore del rendimento a scadenza del titolo in euro
Moto Browniano Geometrico	0,84%	8,04%
Variance-Gamma	1,60%	13,09%
Vasicek	-	0,03%
Vasicek di tipo esponenziale	-	-

Fonte: Elaborazione propria

Pertanto, dall'esame dei risultati ottenuti mediante le simulazioni stocastiche, si osserva che nelle simulazioni basate sul moto Browniano geometrico le probabilità di shortfall, per entrambi i tipi di rendimento a scadenza, risultano inferiori a quelle ottenute nelle simulazioni basate sul processo VG, appartenente alla stessa classe di processi stocastici. Si osserva, inoltre, che nelle simulazioni basate sui processi di tipo Vasicek, caratterizzati dalla proprietà di mean reverting, la probabilità di registrare un rendimento a scadenza superiore a quello del corrispondente titolo denominato in valuta nazionale è pari al 99,97% per il processo di Vasiceck ed è pari al 100% per il processo di Vasicek tipo esponenziale.

Per quanto concerne i percentili riferiti al rendimento a scadenza nelle simulazioni con i diversi processi stocastici, si veda la Tabella 2. In particolare, con riferimento alla colonna del 50° percentile, si evidenzia una tendenza al rafforzamento del dollaro nei confronti dell'euro rispetto al livello d'ingresso in tutti i processi considerati.

Tabella 2. Percentili dei rendimenti a scadenza simulati, per processo generatore dei dati

Processo generatore dei dati	Percentili dei rendimenti a scadenza simulati						
	1°	5°	10°	50°	90°	95°	99°
Moto Browniano Geometrico	0,17%	1,51%	2,26%	4,86%	7,55%	8,32%	9,68%
Variance-Gamma	-0,39%	0,96%	1,69%	4,33%	6,96%	7,78%	9,26%
Vasicek	3,37%	3,67%	3,82%	4,26%	4,67%	4,77%	4,97%
Vasicek di tipo esponenziale	3,17%	3,45%	3,62%	4,20%	4,79%	4,94%	5,26%
Rendimento a scedenza del titolo in euro	1,95%						

Fonte: Elaborazione propria

In conclusione, la nostra analisi effettua una valutazione di un investimento in valuta da parte di un investitore che detiene il titolo con logica di rimborso a scadenza mediante processi stocastici opportunamente scelti.

Dallo studio, alla luce delle simulazioni effettuate, emerge che il differenziale di rendimento tra il titolo denominato in valuta e il titolo denominato in euro può giustificare l'investimento, tenuto conto anche della volatilità del rapporto di cambio.

Tra i processi considerati, il modello di Vasicek ed il modello Vasicek di tipo esponenziale, in virtù della proprietà di mean reverting, appaiono più idonei per la valutazione del rischio di cambio associato all'investimento nella valuta considerata.

Bibliografia

Brigo D., D'Alessandro A., Neugebauer M. and Triki, F. (2007), A stochastic processes toolkit for risk management, available at SSRN: https://ssrn.com/abstract=1109160 or http://dx.doi.org/10.2139/ssrn.1109160.

Engle R.F. (1982), Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica,* 50 (4), 987–1007.

Hull J.C.H. (2009), *Opzioni, futures ed altri derivati*. Pearson Ed., Italia.

Jarque C. M. and Brera A.K. (1987), A Test for Normality of Observations and Regression Residuals, *International Statistical Review,* 55 (2), 163–172.

Ljung G. M. and Box G. E. P. (1978), On a Measure of Lack of Fit in Time Series Models, *Biometrika*, 65 (2), 297–303.

Madan D. B. and Seneta E. (1990), The Variance Gamma (VG) model for share market returns, *Journal of business*, 63 (4), 511-524.

Terborgh G.W. (1926), The Purchasing-Power Parity Theory, *Journal of Political Economy,* 34, 197-208.

Vasicek O. (1977), An equilibrium characterization of the term structure, *Journal of financial economics*, 5 (2), 177-188.

SAPIENZA
UNIVERSITÀ EDITRICE

ANNALI DEL DIPARTIMENTO DI METODI E MODELLI
PER L'ECONOMIA IL TERRITORIO E LA FINANZA
2022 | pp. 119-127

E-ISSN 2611-6634 | ISSN 2385-0825
DOI: 10.13133/2611-6634/1396

Research article

Giuseppe Reale*

THE ROLE OF ASYMMETRY ABOUT INVESTOR PREFERENCES

Abstract

This article is about the role of asymmetry in the distribution of portfolio returns and investors' preferences. It is well known that the skewness of the distribution can play some roles in preferences, but in economic theory this role is usually conflated with the concept of the utility function and, in particular, with expected utility maximization. This perspective seems to us unsatisfactory in two respects. First, the financial intuition about the possibility of accounting for asymmetries is too abstract and difficult for practitioners to grasp. Second, this strategy works only under some implicit conditions such as the existence of third moment; not such a weak assumption for financial returns. Here we propose a different strategy. It considers the comparison between the mean and the median of the distribution of returns. Thus, we obtain a representation that gives us an idea of the possibility of favouring a positive asymmetry and disfavoring a negative one. The main advantage of this representation is that it contains only probabilistic concepts (no utility theory) and is easily understood and communicated by practitioners. Moreover, in this way the existence of a third moment is not necessary, the first one is sufficient.

Keywords: skewness, preferences, moments, distribution.

---

* Sapienza University of Rome, MEMOTEF Department.

1 Introduction

In finance, asset allocation is one of the main problems, and investors' preferences play a crucial role about it. Indeed, the selected portfolio depends on investors' preferences. More precisely, in asset allocation, the distribution of the returns of all potential assets is important, because the distribution of the returns of the selected portfolio depends on the distribution of the assets it contains. Moreover, preferences, at least in general, imply something about the desired distribution of portfolio returns. Therefore, preferences must be linked in some way to the distributional properties of returns.

The importance of measures such as mean and variance of portfolio returns is well known, but asymmetry also plays some roles in the finance literature. However, the intuition about the financial role of mean and variance is trivial, while that about asymmetry is not.

In particular, the desire for the highest possible mean of returns combined with the lowest possible variance is usually considered a necessary condition for an investor to be considered *rational* and *risk averse*. Moreover, this is the starting point for justifying the mean-variance model. This model can be considered as the natural statistical counterpart of the financial concept of return-risk optimization.

Unfortunately, the role of asymmetry is less clear.

Following the paradigm of utility theory, more precisely the criteria of expected utility maximization, the preceding considerations on preferences with respect to mean and variance can be translated into analytical conditions for the *utility function* of the investor [$U(r_p)$]. More precisely, it was shown that the necessary conditions are: positive the first derivative and negative the second derivative. It should be emphasised, however, that this paradigm and its associated conditions are not mandatory to justify a mean-variance model; statistical/financial intuition is sufficient.

In any case, the paradigm of expected utility maximization is the most widespread justification strategy and brings readers to this kind of representation[1]:

$$\mathrm{E}[U(\mathrm{r_p})] = \int_{-\infty}^{+\infty} U(\mathrm{r_p})\,dF(\mathrm{r_p}) = \sum_{n=0}^{\infty} \frac{U^{(n)}(E[\mathrm{r_p}])}{n!} E(\mathrm{r_p} - E[\mathrm{r_p}])^n$$

On the right side, the Taylor expansion series rule around the expected value of r_p (portfolio return) is used. The randomness of r_p leads us to use its distribution [$F(r_p)$], while the Taylor expansion series rule leads us to reduce all relevant features to its moments.

It has been argued that negative asymmetry is an undesirable feature for the distribution of portfolio returns, while positive asymmetry is desirable. In the context of the Taylor expansion series, it was shown that this type of preference is consistent with the positive third derivative of the utility function[2].

Unfortunately, it is not easy to grasp the practical content of this analytical condition. Indeed, sometimes the positive asymmetry is preferred without really understanding why this is a reasonable choice. The most common misunderstandings among

---

[1] For explanations that lead to representations like above, see: Jurczenko and Maillet (2006) section 1.2; Fabozzi et al. (2006) pp.131-137; Jondeau et al. (2007) p. 350.

[2] This result can be generalized. Indeed, it was shown that for risk averse investors skewness should be maximized and kurtosis minimized. In general, odd moments should be maximized and the even ones minimized. Most influential contribution about that, is in: Scott and Horvath (1980).

practitioners arise from the confusion of concepts like: negative asymmetry, lower mean, higher variability and/or risk. For example, there is a misconception that negative asymmetry implies something like a reduction in expected value and/or an increase in the variability of returns. Such reasoning is contradictory because the de-sirability/un-desirability of positive/negative asymmetry should be evaluated *ceteris paribus*, primarily keeping the mean and variance of portfolio returns constant. More problematic discussions are related to the concept of risk and the impact of asymmetry on it. Most of the problems arise from trying to evaluate this effect without even defining the metric of risk.

On the academic side, there is a tendency to consider the undesirability of the negative asymmetry as valid, without dwelling on justifications, and/or to assume the recognition of the positive third derivative of the utility function in question. However, this practice does not seem to be satisfactory on a logical level, because the utility function should encode (predefined) preferences and not suggest/imply them.

In addition, there are other, more technical points. The preceding exposition is not always admissible, since it requires relevant assumptions that are usually passed over in the financial literature. Indeed, the Taylor expansion series rule is reliable only under certain conditions. This rule is also often used in the form of an approximation, but this possibility and the quality of the approximation depend on both the properties of the utility function $U(r_p)$ and the distribution function $F(r_p)$. We cannot analyze them exhaustively here, but some points should be highlighted[3]. First of all, the condition of the positive third derivative mentioned earlier implies that it exists and thus $U(r_p)$ is derivable at least three times. This is not an obvious condition for utility functions and may impose undesirable/unanticipated restrictions on investor preferences. Moreover, $F(r_p)$ must admit moments, at least the same number of moments included in the approximation. Staying with the point discussed here, at least the third moment must exist. This is not such a weak assumption for financial returns, we will come back to this point later.

The following discussion attempts to show in a simple and intuitive way another possible rationale for preferences regarding asymmetry. The idea is to consider only probabilistic concepts, without referring to Utility Theory. In this way, most of the previous problems will be avoided. Moreover, we will only use concepts that can be easily transferable to financial intuition.

2 The concept of asymmetry

In general, the role of asymmetry is to give us clues about the shape of the distribution (indeed, we can refer to asymmetry and kurtosis as indices of shape). Accor-ding to the rule we will use later, we can say that asymmetry tells us something about the shape of the distribution around its median; indeed, under asymmetry, the right and left sides of the distribution are different. However, it must be emphasised that even with asymmetric distributions, the mean remains the most important measure of trend; sometimes this point is not clear enough. A lack of understanding of this fact can lead to misconceptions about the meaning of asymmetry.

It seems like a good idea to talk about a problem right away. In finance, it is common to use terms like asymmetry/skewness and/or third moment and/or index of asymmetry/skewness (by Fisher) as synonyms. This may well be true for the second and third terms, since they are only standardization problems, but not for the first one.

---

[3] For a detailed analysis see Loistl (1976).

Asymmetry in the distribution is a tricky concept and should not be immediately associated with the third moment (or the resulting Fisher's index), even though this is common practice in the finance literature.

2.1 Asymmetry Index of Fisher

To account for asymmetry, several skewness indices have been proposed in the literature; the most commonly used in finance is the Fisher's one. It is based on the central third moment:

$$\mathrm{s}_\mathrm{p}^3 = \mathrm{E}[\mathrm{r}_\mathrm{p} - \upmu_\mathrm{p}]^3$$

where r_p is the portfolio return (the random variable (r.v.) under analysis), whereas $\upmu_\mathrm{p}$ stands for its mean. Skewness index by Fisher is:

$$\frac{\mathrm{s}_\mathrm{p}^3}{\sigma_\mathrm{p}^3}$$

where σ_p is the standard deviation of r_p.

It is so widely and frequently used that, at least in finance, the concept of asymmetry is immediately associated with it without warning; indeed, it is commonly referred to as the "index of asymmetry/skewness". However, it is worth noting that the concept of asymmetry can lead to different measures[4]. In any case, different skewness indices have the property of becoming equal to zero when the analysed distribution is symmetric.

Unfortunately, however, even if a given index is zero, this is no guarantee of the symmetry of the analyzed distribution; this is true even for the Fisher index. It is a necessary but not sufficient condition for proving symmetry. One suggestion would be to compute multiple indices, but this is tedious and, worse, generally insufficient to prove symmetry. In any case, Fisher's skewness index is not an infallible tool for distinguishing between symmetric and asymmetric distributions. Worse, this problem is neither the only nor the most important one for this index.

The main limitation of the Fisher index is that it assumes at least the finiteness of the third moment. Unfortunately, it is worth noting that the existence of third moments for financial returns is not a weak assumption[5]. This issue is related to the problem of fat tails, a quite common and well-known property of financial returns. Therefore, Fisher's skewness index is not a good choice specifically for financial returns. In fact, we can observe that in quantitative finance some distributions are considered useful that have undefined the third moment; for example, α-stable, which never admits the third moment, or *t*-Student, which admits it only when the tail index is greater than 3.

---

[4] For a short review of asymmetry indexes see Piccolo (2010) pp. 157-164.

[5] For a discussion about some problems of the Fisher index of skewness see Kim and White (2004).

2.2 Comparison between mean and median

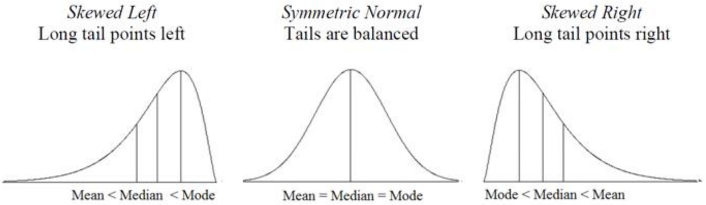
For the discussion of asymmetry, there is a method that focuses on the compari-son among mean, median, and mode. It does not require moments higher than the first one, is therefore more general than the Fisher's skewness index, and it solves at least one of the weaknesses presented earlier.

Some problems are related to multimodal distributions but financial returns usually exhibit best fitting with unimodal alternatives. Moreover, the role of mode is irrelevant in following reasoning. More in particular:

- if mean < median, distribution is negatively (left) skewed
- if mean > median, distribution is positively (right) skewed
- if mean = median, distribution could be symmetric.

It is appropriate to refer on this graph[6]:

Figure 1. Distributions: (negatively) skewed left, symmetric, (positively) skewed right



Source: Doane and Seward (2011)

Unfortunately, it has been shown that even this rule is not infallible[7]; however, it fails only in pathological cases. If we restrict our analysis to continuous and unimodal distributions, most problematic cases cannot arise; fortunately, most relevant cases in finance fit this restriction. This fact leads us to conclude that, at least for financial returns, skewness indices based on the comparison between mean and median are more reliable than those based on the third moment[8].

Shortly, we will propose an argument based on the previous rule. However, to avoid any ambiguity, it is necessary to give a definition of *symmetry in distribution*.

We can say that the distribution of a r.v. X is symmetric if:

$$p(m-x) = p(m+x) \quad \text{for any possible value } x \text{ (of } X\text{)},$$

---

[6] Adapted from Doane and Seward (2011) p. 3.

[7] Indeed in Piccolo (2010, p. 158) it is affirmed that "… *mentre media ≠ mediana implica certamente un'asimmetria nella distribuzione, media = mediana non implica necessariamente una simmetria*", and a related example is given (pp. 163 – 164). In other words, we can say that if the distribution is symmetric then mean = median; however, even if this equality holds, it is not a sure warranty about the symmetry of the distribution. In general, it is not possible to define the concept of symmetry from comparison between mean and median, even if sometimes, maybe for simplicity too, such a definition has been suggested. Indeed, in literature some doubts have been raised about the rule of comparison between mean and median; see von Hippel (2005).

[8] Two indexes based on these quantities are: the Hotelling and Solomon's one and the Yule and Bowley's one. See Piccolo (2010) p. 158.

where $p()$ represents the *density function* of X if X is a continuous r.v., or the *probability function* if X is discrete; m is the median of X.

Moreover, if X is continuous we can add that symmetry implies:

$$F(m - x) = 1 - F(m + x),$$

where $F()$ is the cumulative distribution function (CDF) of X.

An index of asymmetry based directly on this definition is the Bonferroni's one. It is therefore more general than the indices based on moments, however it is rarely used[9].

Moreover, it should be noted that, with the addition of some distributional hy-potheses, the limits of the various skewness indices should be discussed again. For some specified distributions used in quantitative finance, the presence of asymmetry is directly, and indisputably, revealed by some *ad hoc* parameters. An example is the *t-skew distribution*, where the sign of a specific parameter tells us the side of asymmetry, if any; if this parameter is equal to 0, t-skew distribution boils down in the *t-Student* one.

2.3 Representation based on comparison between mean and median

Having clarified the meaning of asymmetry in distribution, we can come back to its relation to investor preferences.

The financial intuition that leads investors to prefer positive (right) skewness and dislike negative (left) skewness is based on Figure 1. Indeed, we can observe that for left-skewed distributions, extremely negative realisations are more likely than the extremely positive ones; while for right-skewed distributions, extremely negative rea-lisations are less likely than the extremely positive ones.

Without loss of generality, let us now consider financial returns as distributions with zero mean: for the right skewed distributions, extreme gains are more likely than extreme losses, while for the left skewed distributions, extreme gains are less likely than extreme losses. For symmetric distributions, gains and losses, extreme or not, are equally likely. This is the financial intuition about the asymmetry of the distribution of returns. However, it is not easy to find a technical explanation for this in the finance literature; we propose here the following one.

From the comparison between mean and median, for negatively skewed distributions, the following must holds:

"mean < median" , therefore $P(X < \mu) < P(X > \mu)$.

Now, it is useful the following decomposition:

$$E[X] = \mu = E[X|X < \mu]P(X < \mu) + E[X|X > \mu]P(X > \mu).$$

Therefore: $\mu - E[X|X < \mu]P(X < \mu) = E[X|X > \mu]P(X > \mu)$;

---

[9] For more information, it is possible to see Piccolo (2010) pp. 159 – 161. A test based on the Bonferroni index is discussed in Mira (1999).

then, considering for simplicity $\mu = 0$, we have that:[10]

$$|E[X|X < 0]P(X < 0)| = |E[X|X > 0]P(X > 0)|.$$

This relation is true in general, regardless equality between $P(X < 0)$ and $P(X > 0)$ (symmetry) or inequality (asymmetry). However, in case of negative asymmetry we have that:

$$|E[X|X < 0]| > |E[X|X > 0]|.$$

This representation helps us to understand that in the case of negative asymmetry, if we look at the absolute values and separate the losses from the gains, the expected losses are greater than the expected gains; in the case of positive asymmetry, the expected losses are smaller than the expected gains; under symmetry, the expected losses and gains are equal.

This proof is about expected gains/losses considered separately, not about extreme events; however, the two concepts are closely related. Indeed, if "expected losses" are larger than "expected gains", it seems sufficient to add some distributive assumptions to ensure that the same result holds for extreme events[11]. In any case, regardless of such assumptions, the result shown above can be taken as an argument that positive asymmetry is desirable and negative asymmetry is undesirable. Such an argument is easily understood by practitioners.

Now, regardless the simplified hypothesis of zero mean, we can find this form:

$$|E[X|X < \mu] - \mu| > |E[X|X > \mu] - \mu|$$

or in more concise terms:

$$\mu - \mu_- > \mu_+ - \mu$$

where $\mu_- = E[X|X < \mu]$ and $\mu_+ = E[X|X > \mu]$. In this case, more care is needed in defining "losses" and "gains"; they should be measured as the difference from the expected return (μ). However, the interpretation remains the previous one. It is also possible to check the following useful equality:

$$\frac{\mu - \mu_-}{\mu_+ - \mu} = \frac{p_+}{p_-}$$

where $p_+ = P(X > \mu)$ and $p_- = P(X < \mu)$.

This means that the relative magnitude of expected losses/gains (asymmetry ratio) depends only on the ratio between the above probabilities. The two ratios are always the same, but they are equal to 1 only under symmetry (>1 in case of negative asymmetry, <1 in case of positive asymmetry).

---

[10] Note that: $| - E[X|X < \mu]P(X < \mu)| = |E[X|X < \mu]P(X < \mu)|$.

[11] We do not face this problem here in a formal way, however we can give some insight. If the distributions are like in Figure 1 (well shaped distributions, then: continuous, unimodal, with monotonic decreasing tails, without truncations, etc.), result showed above for expected values (gain/loss) seems generalizable to extreme events too. Graphs suggest this conjecture.

This ratio can be considered an index of asymmetry. However, in order to achieve a more conventional reading the following simple transformation is needed:

$$s = 1 - \frac{p_+}{p_-};$$

if the distribution under analysis is left skewed $s < 0$, if it is right skewed $s > 0$, if it is symmetric $s = 0$.

Moreover, it is worth noting that under positive asymmetry it holds:
$p_+ < p_-$ and $\mu - \mu_- < \mu_+ - \mu$ (probability of loss greater but expected loss lower).

This situation, that should be the desired one, at first glance can seem strange for practitioners, due to $p_+ < p_-$ (probability of gains smaller than that of losses). However, it must hold if the three kinds of distribution (negatively skewed, positively skewed, symmetric) must share the same mean, therefore the same redditivity[12]. Indeed, the desirability of positively skewed distributions is based only on the reduction of the relative dimension of expected/extreme losses, in comparison with expec-ted/extreme gains, not on the general redditivity.

3 Conclusions

The fact that for securities/portfolios a positive asymmetry of the return distribution is better than a negative one is generally accepted in finance.

However, the reasons for this fact are sometimes misunderstood or simply igno-red by practitioners. On the academic side, the most common explanation is based on quite abstract concepts such as analytical conditions on utility functions. These ratio-nales are not easily communicated to practitioners, and their financial intuition may remain obscure. Worse, this rationale requires some implicit conditions such as the existence of third moments; not such a weak assumption for financial return distribu-tions.

The explanation presented here does not involve utility theory and related analytical conditions, but only probabilistic ones. The main results are directly transferable to financial intuition and therefore easily communicated to practitioners. Moreover, this reasoning works independently on the existence of moments greater than the first. It is therefore more general than those involving the third moments and allows one to avoid associated difficulties.

References

Doane D.P. and Seward L.E. (2011), Measuring Skewness: A Forgotten Statistic?, *Journal of Statistics Education,* 19 (2), available at http://jse.amstat.org/v19n2/doane.pdf.

Fabozzi F.J., Focardi S.M. and Kolm P.N. (2006), *Financial Modeling of the Equity Market: From CAPM to Cointegration.* John Wiley & Sons, Hoboken, New Jersey.

---

[12] It may be noted that we did not say anything about the variance of returns. It is sufficient so to state that the three distributions we discussed (negatively skewed, positively skewed, symmetric) explicitly share the same mean but may also share the same variance.

Jondeau E., Poon S-H. and Rockinger M. (2007), *Financial Modeling Under Non-Gaussian Distributions.* Springer Finance, Springer-Verlag, London, https://doi.org/10.1007/978-1-84628-696-4.

Jurczenko E. and Maillet B. Eds. (2006), *Multi-moment Asset Allocation and Pricing Models.* John Wiley and Sons, Chichester, England, DOI:10.1002/9781119201830.

Kim T.H. and White H. (2004), On More Robust Estimation of Skewness and Kurtosis, *Finance Research Letters*, 1 (1), 56–73, https://doi.org/10.1016/S1544-6123(03)00003-5.

Loistl O. (1976), The Erroneous Approximation of Expected Utility by Means of a Taylor's Series Expansion: Analytic and Computational Results, *The American Economic Review*, 66 (5), (Dec., 1976), 904- 910.

Mira A. (1999), Distribution-free test for symmetry based on Bonferroni's measure, *Journal of Applied Statistics,* 26 (8), 959- 972, https://doi.org/10.1080/02664769921963.

Piccolo D. (2010), *Statistica* (terza edizione). Il Mulino, Bologna.

Scott R.C. and Horvath P.H. (1980), On the Direction of Preference for Moments of Higher Order Than Variance, *The Journal of Finance,* 35 (4), 915–919, https://doi.org/10.1111/j.1540-6261.1980.tb03509.x.

von Hippel P.T. (2005), Mean, Median, and Skew: Correcting a Textbook Rule, *Journal of Statistics Education,* 13 (2), https://doi.org/10.1080/10691898.2005.11910556.

NOTES AND DISCUSSIONS

SAPIENZA
UNIVERSITÀ EDITRICE
ANNALI DEL DIPARTIMENTO DI METODI E MODELLI PER L'ECONOMIA IL TERRITORIO E LA FINANZA
2022 | pp. 131-137

E-ISSN 2611-6634 | ISSN 2385-0825
DOI: 10.13133/2611-6634/1325

Notes and discussions

Marco Brogna*

L'IMPATTO DELLA PANDEMIA SUL TURISMO IN ITALIA.
ALCUNE RIFLESSIONI SULL'OCCUPAZIONE

Abstract

The impact that the COVID-19 pandemic has produced on tourism has been, and in part still is, dramatic from a quantitative point of view and, in many ways, not entirely predictable from the point of view of qualitative and structural transformations. Much specialist literature has been produced on the subject practically in real time and the debate, especially from the point of view of the tourism development model of the future, is still open. In this note, we try to contribute to this discussion through some observations made on the dimension of tourist occupation in Italy. It seems, in fact, that the pandemic has highlighted the many factors of structural weakness in the sector, and that the message to be received is decidedly that of a significant change in the objectives and methods of tourism development.

Keywords: tourism, pandemic, employment, tourism policies.

---

* Sapienza University of Rome

1 Introduzione

Il ruolo del turismo come volano delle economie regionali, nonché le criticità strutturali del settore, non sono certamente temi nuovi nel dibattito accademico e politico, spesso al centro di accesi confronti tra rappresentanti del mondo delle imprese, portatori di interesse e istituzioni nazionali e sovranazionali (Sinclair, 1998; Lew, 2011; Cárdenas-García et al., 2015).

La pandemia da COVID-19 sembra aver puntato ancor meglio i riflettori sulla posizione fondamentale che il turismo riveste, in alcune regioni in modo particolare, come fonte di occupazione e reddito (in Italia, ad esempio, nel 2019 al turismo va imputato il 13% del PIL) e, d'altra parte, sui molti elementi di debolezza strutturale sui quali sembra ormai ineludibile intervenire in modo deciso e (Skare et al., 2021; Collins Kreiner and Ram, 2021).

Uno di questi elementi critici è rappresentato dal ruolo del turismo come fonte occupazionale, considerando che si tratta di uno dei settori produttivi più labour-intensive. Quali saranno gli effetti sull'occupazione -in termini macroeconomici ma anche considerando le diverse nuove configurazioni della struttura interna- e sulla sua articolazione territoriale, della riduzione del volume dei flussi turistici? Che tipo di profili occupazionali emergeranno quando gli effetti devastanti del COVID (WORLD TOURISM ORGANIZATION, 2021a) saranno cessati e il settore potrà ripartire? Il dibattito è molto vivace in letteratura, e l'osservazione del caso italiano sembra poter contribuire alla costruzione di un modello turistico più attento alla dimensione della qualità dell'occupazione e della tutela delle dimensioni ambientali e sociali che ne sono costitutive (Assaf and Scuderi, 2020; Hall et al., 2020a and 2020b).

In questa nota si intende mettere in luce alcuni fattori critici del turismo in Italia, in particolare nella componente dell'occupazione e della struttura professionale nel settore, giungendo ad alcune proposte di intervento che la pandemia sembra aver reso improcrastinabili.

2 La struttura occupazionale del turismo in Italia

La struttura dell'occupazione turistica in Italia è decisamente caratterizzata da mansioni di basso profilo anche se, in anni recenti, si registrano percentuali crescenti di professioni con elevati livelli di formazione e specializzazione, in particolare nelle varie dimensioni dell'ICT.

Altra nota caratteristica dell'occupazione turistica in Italia è l'incidenza elevata dei contratti a termine (secondo l'ISTAT il 26,2% contro il 13,1% della media nazionale) e l'impiego frequente e capillare di lavoratori part-time (28,7% a fronte del 19% nazionale). Fattori che, se a lungo hanno rappresentato per l'impresa un punto di forza, contribuendo a rendere efficiente e particolarmente elastica l'offerta rispetto ad una domanda in continua e rapida evoluzione, vengono ormai considerati giustamente come un aspetto critico del sistema turistico.

Ancora, nell'assetto dell'occupazione turistica in Italia si nota una particolare concentrazione in alcuni segmenti: dieci figure professionali polarizzano il 74% degli occupati; di queste, le prime cinque rientrano in modo pressoché esclusivo nel comparto della ristorazione (baristi, camerieri, cuochi, esercenti nelle attività di ristorazione, addetti alla preparazione, cottura e distribuzione di cibi), attività che ha sofferto più delle altre le chiusure forzate e le restrizioni operative decretate in tempi di pandemia.

La complessità di questa articolazione interna all'occupazione turistica va osservata nella sua distribuzione geografica, considerando che i lavoratori del settore turistico sono maggiormente presenti nelle regioni del Centro-Sud (52,9% rispetto al 47,8% del totale occupati), si caratterizzano per una elevata componente femminile (45,4% contro il 42,3% nazionale) e, soprattutto, per una maggiore quota di giovani tra i 15 e i 34 anni (38,3% a fronte del 22,1% nazionale). Era inevitabile che l'impatto della pandemia sulle imprese turistiche producesse una significativa perdita di posti di lavoro e che questo colpisse regioni già in difficoltà, soprattutto indebolendo le possibilità di occupazione di popolazioni giovani e di sesso femminile, ridimensionando i risultati attesi dai vari interventi di politica per lo sviluppo turistico messi in campo dagli ultimi cinque governi nazionali.

Se, infine, all'impatto diretto si aggiunge quello prodotto sull'indotto turistico, la crisi occupazionale si preannuncia ben più grave. L'impatto è comprensibile se si considera come, secondo alcune stime, il turismo è in grado di generare un posto di lavoro nel settore secondario ogni quattro posti di lavoro creati nel settore stesso (ISTAT, 2020).

L'entità della perdita di posti di lavoro causata dalla crisi non è ancora del tutto manifesta, anche perché la fase pandemica non è ancora conclusa. Molti governi, l'Italia tra questi, hanno protetto i lavoratori dall'impatto della pandemia con strumenti congiunturali di breve durata. Si è trattato di pura sopravvivenza economica, di strumenti di sostegno sociale a famiglie non più in grado di produrre reddito.

Tuttavia, si stima che circa 100 milioni di posti di lavoro siano attualmente a rischio nel mondo (Marques Santos et al., 2021) mentre in Italia il 14% delle aziende del comparto turismo e servizi di ospitalità è a un passo dal fallimento (il 10% è già fallito tra 2020 e primo semestre 2021), con quasi 100mila posti di lavoro ad altissimo rischio (CERVED, 2020). Si tratta soprattutto di dipendenti di aziende di piccole e medie dimensioni con una forte prevalenza di microimprese (meno di 10 addetti). Proprio quelle microimprese spesso gestite da giovani, donne e lavoratori "informali" caratterizzati dal rivolgere la propria domanda di lavoro verso le medesime classi sociali.

3 Il quadro regionale

Dato l'insieme di peculiarità appena osservate, e considerando come altri settori economici, pur vivendo la medesima crisi pandemica, non hanno registrato i danni strutturali tipici del turismo (tutte le stime sulla possibile ripresa del turismo, per quanto ottimistiche, vedono come anno di riavvio il 2023 (WORLD TOURISM ORGANIZATION, 2021a), la caduta dell'occupazione in termini quantitativi potrebbe portare con sé anche una dannosa perdita di competenze nel settore, poiché molti lavoratori in uscita dal turismo si ricicleranno, in modo spontaneo, in altri comparti.

Un'emorragia di forza lavoro che coinvolgerebbe in misura prioritaria le figure più qualificate e a maggior remunerazione, professionalità troppo onerose per essere trattenute, ma al tempo stesso le uniche in grado di garantire le necessarie abilità per ripartire dopo la crisi. Inoltre si potrebbe innescare un fenomeno assolutamente negativo in termini di riduzione delle prospettive occupazionali nel settore, spingendo la ricerca di occupazione verso altre mete e altri obiettivi, indebolendo ulteriormente un settore che da subito dovrebbe schierare le forze migliori in termini di professionalità, di impegno e di idee, se vuole tornare a competere con le altre destinazioni top nel mondo, riattivando quel fenomeno di crescita e sviluppo turistico raggiunto con successo nel 2019 (CENSIS, 2021).

Purtroppo, sul fronte della performance delle imprese turistiche si registrano notizie poco incoraggianti. Una recente indagine ISTAT sull'impatto dell'emergenza sanitaria ed economica ha analizzato un campione di oltre 177mila aziende turistiche, pari al 42,9% delle imprese turistiche italiane, al 72,3% degli addetti in esse impiegati e al 79,8% del valore aggiunto (ISTAT, 2021a)[1]. Il risultato mostra come il settore turistico si confermi fra i più colpiti dall'emergenza sanitaria. Nei mesi del primo lockdown nazionale (marzo-aprile 2020) il 38,2% delle imprese turistiche ha dichiarato di non avere registrato fatturato e più della metà (50,9%) ha dichiarato che il valore delle vendite è stato più che dimezzato rispetto agli stessi mesi del 2019. Tra giugno e ottobre 2020 l'entità della perdita si è attenuata, ma si è confermata ampia la quota di imprese colpite: quasi 8 su 10 (il 79,6%) hanno dichiarato una flessione del fatturato superiore al 50% (il 27,4%) o compresa tra il 25-50% (il 52,2%), rispetto a una media del complesso dei settori più contenuta.

Ancora, secondo l'indagine ISTAT, la crisi sembra rivelare chiaramente delle differenze regionali di performance, anche in termini di conseguenti perdite di posti di lavoro. Tra giugno e ottobre del 2020 il Mezzogiorno ha registrato la quota più elevata di imprese con calo molto elevato del fatturato: l'83,1% ha dichiarato una diminuzione superiore al 50% (32,7% media nazionale), rispetto all'80,8% registrato al Centro e nel Nord-Ovest e al 74,4% nel Nord-Est.

È anche vero che ragionare in termini di "tre Italie" può risultare riduttivo: l'impatto economico della pandemia sui territori è stato infatti molto eterogeneo. Alcune regioni sono riuscite a contenere gli impatti negativi facendo leva su altri settori economici; altre, a minor incidenza del settore, hanno ovviamente sofferto meno per la crisi turistica. Sulla base dei risultati delle indagini sugli effetti della crisi da COVID-19, in undici regioni almeno la metà delle imprese presenta un rischio alto o medio-alto (riduzione di fatturato, seri rischi operativi e assenza di strategie di reazione alla crisi); di queste, sette sono nel Mezzogiorno (Campania, Abruzzo, Basilicata, Calabria, Sicilia, Sardegna, Puglia), una al Nord (Provincia autonoma di Bolzano) e tre nel Centro Italia (Lazio, Umbria e Toscana).

Pertanto, ad oggi circa un terzo degli addetti totali (32,6%) è impiegato in imprese che registrano un livello di rischio catalogato come alto o medio-alto, con percentuali

---

[1] L'ISTAT ha elaborato cinque classi di imprese, in base alle strategie adottate per affrontare la crisi: "Imprese statiche in crisi", che stanno subendo pesantemente l'impatto dell'emergenza sanitaria senza riuscire a mettere in campo strategie ben definite; "Imprese statiche resilienti", che nonostante una carenza di azioni hanno registrato limitati impatti negativi; "Imprese proattive in sofferenza", duramente colpite dalla crisi che hanno intrapreso specifiche iniziative; "Imprese proattive in espansione", colpite in modo marginale dalla crisi che proseguono sui loro precedenti piani di sviluppo; "Imprese proattive avanzate", che nel periodo 2020-2021 hanno aumentato gli investimenti. La distribuzione delle imprese turistiche all'interno delle cinque classi testimonia l'assenza di un adeguato dinamismo settoriale: il 63,3% delle imprese è statico e rientra nella prima e seconda classe. Addirittura, il 50% è composto da aziende statiche in crisi, a fronte del 28,6% di imprese statiche nel settore industria e servizi. Il 36,7% di aziende turistiche è proattivo, al fronte del 71,4% di industria e servizi nel suo complesso. Se è vero che il turismo è il settore più colpito dalla pandemia è allo stesso tempo il più statico, incapace di avviare strategie di contenimento delle perdite e di rilancio dei propri risultati.

che in alcune regioni superano il 40%, ed anche in questo caso, con marcati divari regionali: sette sono nel Mezzogiorno (Basilicata, Calabria, Abruzzo, Sardegna, Molise, Sicilia e Campania), una nel Centro (Umbria) e una nel Nord (Valle d'Aosta).

Ancora Istat (ISTAT, 2021b) ha elaborato un indicatore territoriale di "rischio combinato" (sintesi del rischio per imprese e addetti), al fine di porre in evidenza quanto la crisi sia destinata ad accentuare il divario tra le aree geografiche: delle sei regioni il cui tessuto produttivo risulta ad alto rischio, cinque appartengono al Mezzogiorno (Abruzzo, Basilicata, Calabria, Campania e Sardegna) e una al Centro (Umbria) mentre le sei a rischio basso sono tutte nell'Italia settentrionale (Piemonte, Liguria, Lombardia, Emilia-Romagna, Friuli-Venezia Giulia, Provincia autonoma di Trento). La situazione di estrema fragilità del comparto occupazionale turistico è purtroppo confermata dalla ricerca ISTAT sui sistemi locali del lavoro (ISTAT, 2021c), nel corso della quale si è calcolato un "indice di rischio territoriale" sulla base della collocazione delle stesse imprese nei sistemi locali. Anche in questo caso è emersa una chiara dicotomia tra Nord e Sud, con il primo caratterizzato da un sistema di imprese meno fragile e il secondo da una esposizione al rischio significativamente maggiore.

Come si diceva, tuttavia, il problema non è solo e tutto meridionale. Livelli elevati di vulnerabilità si riscontrano anche in aree del Centro (Toscana, Lazio e Umbria) e del Nord (Valle d'Aosta e Provincia autonoma di Bolzano), a conferma che all'interno di un quadro generale quantomeno di stagnazione economica, la geografia della vulnerabilità risulta molto differenziata e testimonia un generale problema strutturale del turismo nazionale.

Infatti, anche nelle regioni settentrionali emergono numerose realtà locali fragili, soprattutto all'interno di sistemi locali del lavoro a forte specializzazione turistica (Susa, Courmayeur, Livigno, Ponte di Legno, San Candido, Pinzolo, Jesolo, Finale Ligure, Sestri Levante, Cesenatico).

Nelle regioni del Centro, le aree a maggiore fragilità sono individuabili soprattutto nelle zone agricole e turistiche della Toscana (Monte Argentario, Orbetello, Montalcino, Portoferraio, ad esempio), dell'alto Lazio (Acquapendente, Civita Castellana) e in alcune zone dell'Umbria (Cascia, Norcia) e del basso Lazio (Sabaudia, Gaeta, Terracina). I sistemi locali distrettuali presentano invece una minore rischiosità, anche grazie alla presenza di occupazione in settori manifatturieri o ad essi collegati. È il caso, in particolare, dei centri di Arezzo e Lucca, specializzati rispettivamente nell'oreficeria e strumenti musicali, e nell'industria cartotecnica.

Nel Mezzogiorno, mentre Abruzzo e Molise hanno un numero limitato di sistemi locali in difficoltà, altri territori, soprattutto quelli più specializzati nel turismo, sono in sofferenza: dai sistemi locali di Capri, Ischia e Amalfi in Campania, a quelli di San Giovanni Rotondo, Fasano, Ostuni, Gallipoli in Puglia, a Maratea e Policoro in Basilicata, a Praia a Mare, Tropea e Cirò Marina in Calabria. Lo stesso in Sicilia: Gela, Sciacca, Licata, Noto fra le zone turistiche, Castelvetrano e Pachino fra quelle più agricole. La Sardegna mostra una realtà peculiare in cui la distinzione tra zone interne (più penalizzate) e costiere (in alcuni casi meno svantaggiate) sembra essere sostituita da una contrapposizione tra sistemi locali urbani (Cagliari e Sassari), relativamente solidi, e sistemi non urbani, in difficoltà.

4 Alcune considerazioni conclusive

Dal quadro tratteggiato, dagli spunti che si è provato a dare, emerge tutta la complessità della situazione occupazionale del turismo in Italia, e molte delle criticità strutturali del comparto e degli effetti prodotti dalla pandemia.

Gli scenari che si prefigurano non sono ovviamente rosei, considerando che secondo il WORLD TOURISM ORGANIZATION (2021b) l'economia legata al turismo avrà bisogno di molto tempo per recuperare le posizioni perdute e non è comunque affatto sicuro che ci riuscirà. In questo quadro, l'Italia non potrà accontentarsi di recuperare le posizioni perdute.

Il sostegno economico al turismo non dovrà essere sicuramente l'unica misura da mettere in campo, come ribadito da molti esperti del settore. Certamente condivisibile il bisogno di passare dalla fase delle sovvenzioni a quelle degli incentivi alla ripresa, così come misure finanziarie alle imprese e alle destinazioni turistiche, o a misure di tutela per le imprese affinché le più fragili non vengano assorbite dai colossi transnazionali del turismo (Assaf and Scuderi, 2020).

Tuttavia, non è possibile puntare ad un ritorno al business as usual, né per la competitività del settore né per la sostenibilità dello sviluppo turistico. L'innovazione e la formazione devono essere le misure di punta per il sostegno all'occupazione turistica e per mantenere in vita un comparto essenziale per l'economia italiana. Al di là dei soliti dati quantitativi sui flussi, vanno considerate altre dimensioni dello sviluppo turistico, anche a partire dal sostegno della qualità occupazionale e dal contenimento di un rischio di fuga della forza lavoro verso altri settori, o ancor peggio verso altri paesi, che è presente e crescente.

La struttura occupazionale osservata, con i suoi caratteri peculiari al comparto e all'Italia, mette il nostro paese in una situazione molto delicata. Inutile dire che è necessario ora più che mai puntare sulla crescita formativa, sull'incremento delle competenze specializzate nel settore, su quelle conoscenze e su quelle capacità operative necessarie per tornare ad essere competitivi nel mercato globale. Utile, invece, ribadire ancora una volta l'esigenza di lavorare in modo transdisciplinare per individuare le leve più appropriate a trattenere e a sviluppare queste professionalità, attivando finalmente il giusto mix di interventi pubblico-privati per la qualificazione di un settore che continua a dimostrarsi, nonostante tutto, un grande motore di sviluppo per l'Italia.

References

Assaf A., Scuderi R. (2020), COVID-19 and the recovery of the tourism industry, *Tourism Economics*, 26(5), 731-733.

Cárdenas-García P.J., Sánchez-Rivero M., Pulido-Fernández J.I. (2015), Does tourism growth influence economic development?, *Journal of Travel Research*, 54(2), 206-221.

CENSIS (2021), Italia sotto sforzo. Diario della transizione 2020/21. Reinventare il turismo dopo la tempesta perfetta, available at https://www.censis.it/sites/default/files/downloads/Diario%20della%20transizione_4_2021.pdf [10 febbraio, 2022].

CERVED (2020), L'impatto del Covid-19 sulla filiera turismo e trasporti, 2020. Covid-19: Risposte di Policy per le PMI, available at https://know.cerved.com/imprese-mercati/le-pmi-italiane-alla-prova-del-covid/ [8 febbraio, 2022].

Collins Kreiner N., Ram Y. (2021), National tourism strategies during the Covid-19 pandemic, *Annals of Tourism Research*, 86, 103076.

Hall C.M., Scott D., Gössling S. (2020a), Pandemics, transformations and tourism: be careful what you wish for, *Tourism Geographies*, 22(3), 577-598.

Hall C.M., Scott D., Gössling S. (2020b), Pandemics, transformations and tourism, *Journal of Sustainable Tourism*, 29(1), 1-20.

ISTAT (2020), Movimento Turistico in Italia (gennaio-settembre 2020), available at https://www.istat.it/it/files/2020/12/REPORT_TURISMO_2020.pdf [5 febbraio, 2022].

ISTAT (2021a), I profili strategici e operativi delle imprese italiane nella crisi generata dal Covid-19, available at https://www.istat.it/it/archivio/252396 [2 febbraio, 2022].

ISTAT (2021b), III trimestre 2020: Fatturato dei servizi. Statistiche Flash, available at https://www.istat.it/it/files//2020/11/fs_0320.pdf [2 febbraio, 2022].

ISTAT (2021c), Rapporto sulla competitività dei settori produttivi, available at https://www.istat.it/it/files//2021/04/Rapporto-Competitivit%C3%A0_2021.pdf [10 gennaio, 2022].

Lew A. A. (2011), Tourism's Role in the Global Economy, *Tourism Geographies*, 13(1), 148-151.

Marques Santos A., Madrid C., Haegeman K., Rainoldi A. (2021), Behavioural changes in tourism in times of Covid-19, Publications Office of the European Union, ISBN: 978-92-76-20401-5.

Sinclair M.T. (1998), Tourism and Economic Development: A Survey, Journal of Development Studies, 34(5), 1-51.

Skare M., Soriano D.R., Porada M. (2021), Impact of COVID-19 on the travel and tourism industry, *Technological Forecasting & Social Change*, 163, 120469.

WORLD TOURISM ORGANIZATION (2021a), Conceptual Guidance on Tourism Statistics in the COVID-19 Context, UNWTO, Madrid.

WORLD TOURISM ORGANIZATION (2021b), Investment Guidelines – Strategies to Safe-guard Tourism Investments during COVID-19, UNWTO, Madrid.

SAPIENZA
UNIVERSITÀ EDITRICE

ANNALI DEL DIPARTIMENTO DI METODI E MODELLI
PER L'ECONOMIA IL TERRITORIO E LA FINANZA
2022 | pp. 139-149

E-ISSN 2611-6634 | ISSN 2385-0825
DOI: 10.13133/2611-6634/1397

Notes and discussions

Carla Ventre*

EUROPEAN MIGRATION POLICIES BETWEEN RESTRICTIONS AND FEW RIGHTS. THE CASE OF THE BALKAN ROUTE

Abstract

The rise of new migratory routes in the world has represented a response to the unsuccessful strategies adopted by States in the field of people mobility. The Balkan route, in Eastern Europe, came out from the migration crisis in 2015 and continues until today. The main goal of this note is to provide a detailed literature regarding the policies adopted by the States to manage migration along Balkan countries. It is also the aim of this paper to show how policies have produced changes in the route, violating the fundamental human rights instead of reducing the number of people trying to cross the borders to reach Europe but rather creating categories of marginalized people in marginalized places.

Keywords: Migration, Balkan route, European migration policies, the game, pushbacks.

---

* Department of Social and Economic Sciences - Sapienza University of Rome

1 The Balkan route

Migration is one of the burning issues of contemporary societies. In a world in which distances no longer seem to exist and where goods are exchanged with extreme ease, some people still find it greatly difficult to move from one country to another. This article focuses on these people, who often come from the poorest countries of the world affected by wars, internal conflicts, poverty and the effects of climate changes. In absence of legal pathways, migrants travel along routes that change with the political process. On the basis of the high principles and values that make Europe unique in the world, the article aims to raise awareness regarding what happens along the European borders and regarding how, particularly in the last years, the European migration policies are failing in their goals at the expense of human rights.

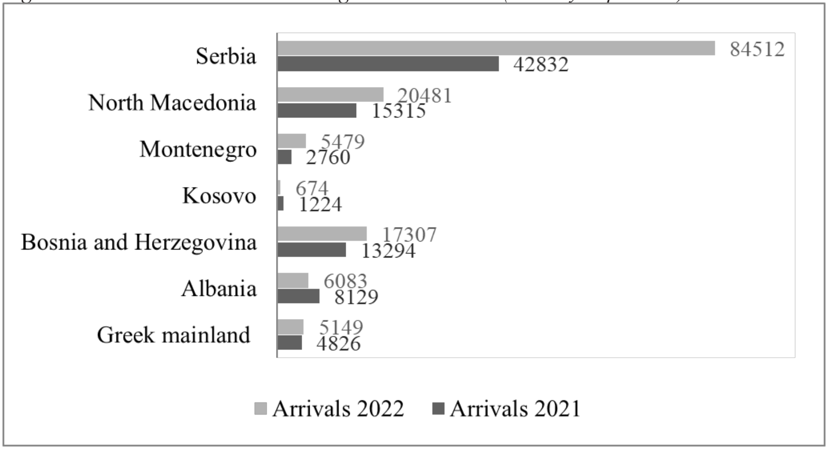
The area of interest for this study is the Balkan one regarding which, because of the complex situation widespread in the area, there is a gap in literature concerning the dynamics working there. This note proposes an analysis of the transformation processes of the Balkan route until today and how the migration policies in the countries have changed in line with the interests of Europe. Moreover, the article highlights what are socio-political factors that have an impact in the changes. Through the data collected by the Aegean Boat Report, which monitors issues related to people movements in the Aegean Sea, and a comparative analysis with data by other official international organisations, it was possible to show the negative effects that European policies have caused along the borders.

In public debate, Alan Kurdi's history (Ibrahim, 2018), a child died on a Turkish coast in an attempt to reach Europe, shifted the attention from the shipwrecks in the central Mediterranean to the long crossing of the Balkan route. Several people crossed the Balkans after landing in Greece from Turkey, but many have died on this route or been deprived of the most basic rights along the same borders.

Europe decided to implement a more humanitarian approach: in April 2015, a formalized corridor was established by the States for the transit of people. The corridor facilitated the movement of people from Greece to Central and Northern Europe crossing Serbia, North Macedonia, Croatia and Slovenia. However, the corridor was opened for only one year and it remains a historical event in recent EU history (Abikova and Wojciech, 2021). The rules of the corridor and its pathways were established thanks to the cooperation among the countries along the route by agreements fostered by the EU institutions, until the strong position of Hungary and Austria against the corridor drove the process of closing the corridor. At the same time, the European Commission led to a new process of Europeanization of migration policies and its externalization to non-European countries. The border was closed and thousands of migrants were stuck in the Balkan countries. Since then, Europe has financed Turkey to stop migrants from reaching Greece, the first European country of the route. The European institutions have established reception camps along the Balkan route for those migrants that despite the closure of borders are able to cross the borders relying on smugglers and traffickers. Moreover, this new form of cooperation among Balkan countries and European institutions allows formal expulsions at borders, the so-called *pushbacks*, during which migrants are denied applying for their asylum requests and suffer from subsequent violence and abuses by border police.

The data collected on the arrivals at the Greek mainland (UNHCR OPERATIONAL DATA PORTAL, 2021b) and the other countries of the Balkan route suggest that the policies adopted by the European States cause a change of routes towards other countries (Figure 1).

Figure 1. Arrivals at the countries along the Balkan route (January-September)



Source: UNHCR OPERATIONAL DATA PORTAL (2022 and 2021a)

Moreover, in the last year, one of the most relevant events to demonstrate how migration is strictly linked to the show of strength among governments is the migration crisis along Poland (EU) – Belarus border. The crisis started when, in the summer 2021, Belarus increased the number of flight connections from the Middle East giving a visa-free travel for 30 days. Migrants arrived mainly from Iraq, Afghanistan, Syria and Turkey with the hope to reach Europe through the neighbouring Poland, Lithuania and Latvia. We have not much information regarding the reasons that have driven Belarusian President to promote these actions. Probably because of European sanctions inflicted on Belarus for the violations committed against the political protests in the country or simply by putting pressure on European States. In any case, even today, people on the move continue to be stuck in a forested border region in need of medical assistance, where they face freezing temperatures without food, water, shelter and warm clothes. Migrants suffer violence from the Belarussian army on the one hand, and from the Polish army on the other. International organisations such as Médecins Sans Frontières denounce the EU pushback policies and restricted access for aid organisations in forested border regions in which 21 people have already lost their lives (Médecins Sans Frontières, 2022). So, along the borders of “Fortress Europe” the dynamics do not change.

1.1 How the Balkan route is changing

In the history of European migration, the Balkan corridor remains the most solidarity response to the migration crisis adopted by European countries to help people to move into frontiers. In early summer 2015, the formalized corridor enabled refugees to cross the Balkans from Greece to Western Europe within two or three days by trains and buses. This policy guaranteed safe passages for migrants, who did not rely on smugglers or risked their lives in dangerous circumstances (Stojić Mitrović et al., 2016). The corridor was opened in different modalities and thanks to a strong cooperation among European countries and non-European countries, until March 2016.

As a result, 1.5 million people crossed the Balkans led by police on the move to follow the corridor's direction: from Greece to Central and Northern Europe crossing Serbia, North Macedonia, Croatia and Slovenia (El-Shaarawi and Razsa, 2018). From Turkey to Aegean Islands migrants were stopped in the hotspots, instituted in May 2015 in the European Agenda on Migration (European Commission, 2015a), with the aim to help the Member States bordering with non-EU countries, to identify, register and fingerprint incoming people. Five hotspots were created in Greece, on the islands of Chios, Kos, Leros, Lesbos and Samos.

The registration was valid for 30 days and it enabled migrants to travel the mainland in a legal way.

Some countries, such as Serbia, benefited from the increasing number of migrants in their territory for promoting their accession process to the EU. On the contrary, in Greece there were several disputes, a clear sign of the unsatisfactory response of Europe in managing migration. The strong position of Hungary and then also of Austria against the corridor drove the process of closing. The result was that more than 75,500 migrants were stuck in transit countries, unable to move (Abikova and Wojciech, 2021). Another strategy to cut access to the formalized corridor was achieved by adopting the EU-Turkey deal signed on 18 March 2016. The migration became a security issue, a danger for public order, cultural identity, domestic and labour market stability. Although the European policies were influenced by humanitarian principles of governments based on alleviation of suffering and preservation of life, the border was closed. Despite these, people have continued travelling using more dangerous roads (Hameršak et al., 2020).

The newly adopted policies have changed the migration pathway and new countries take part in the Balkan route in different ways. The closure of the corridor and stricter border controls caused a dramatic transformation of the Balkan dynamics in recent years. The more restrictions are imposed to cross the countries, the more dangerous branches rise in the route. So, along the Balkan route migrants, who haven't possibility of reaching Europe in a safe way, try to arrive in Europe relying on smugglers and traffickers.

The biggest problem for migrants on the move is the accommodation. There are few accommodation centres with a limited capacity in transit countries. In 2018 the EU decided to collaborate directly with the International Organisation for Migration (IOM) and with United Nation agencies as partners instead of State institutions, in managing migration. Despite the great number of donations, they failed to create dignified and human living conditions, only offering places where people feel insecure and where psychological support and health care are limited. People, to fulfil the legal obligation, have to register an official address within a few days, but this process became impossible because of the lack of accommodation and the currently temporary solutions. The result is that a large number of people without personal documents and a regulated stay are considered illegal migrants and live in precarious conditions (Ahmetašević, 2020).

1.2 What happens along Balkan route?

People on move are often young men between 17 and 30 years old, but also women and children. They face the so called *game*: people struggle to walk for days through mountains and it happens that they make arrangements with smugglers to be transported by car from the country they are leaving to Europe, they wait for days in highly hidden places fearing to be discovered by the police and turned away (Sapoch and Augustova, 2020). These remote locations are known as *squat* and *jungle camps*. The squats are abandoned and weathered houses, in peripheral and hilly areas, used as night shelter, with no water and electricity and limited hygienic conditions. When the squats are full, people on the move settle in the forests, creating so-called jungle camps, with the same degraded conditions. When the police discover the migrants push them towards the borders of neighbouring countries.

Pushback procedures are part of an interconnected network of actors during which several people suffer violence and physically attacks by border police (Sapoch and Augustova, 2020). Denial to asylum is one of pushback practice that migrants suffer along borders. The asylum requests are often ignored or shut down by verbal and physical attacks. Several organisations record cases of robbery and damage to private belongings. Moreover, Médecins Sans Frontières denounce the unnecessary force used by border guards on their patients along the route: black eyes, sprained ankles, broken legs and arms, pepper spray burns, and foot long bruises indicate the extreme violence exercised (Médecins Sans Frontières, 2018).

2 The role of EU

On 25 October 2015, leaders representing Albania, Austria, Bulgaria, Croatia, the former Yugoslav Republic of Macedonia, Germany, Greece, Hungary, Romania, Serbia and Slovenia met in Brussels with the European Commission.

During the Meeting on the Western Balkans Migration Route the country leaders agreed on a 17-point Plan of Action (European Commission, 2015b). Namely, they agreed to cooperate with each other along the route and decided on pragmatic operational measures that can be implemented to tackle the refugee crisis in the region.

The States set up the operational measures to allow a permanent exchange of information, to control movement of people along the route, to discourage the movement of people to the border of another country or region without informing neighbouring countries.

With regard to support refugees, States increase the capacity to provide temporary shelter, food, health, water and sanitation to all in need.

In the management of migration States ensure a full capacity to register arrivals with the use of biometric data through a close cooperation with EU agencies. Frontex and other EU agencies step up efforts to return migrants not in need of international protection and to cooperate on readmission. Leaders also approve an increase of efforts to manage borders, including: implementing the EU-Turkey Action Plan (European Commission, 2015b); using the potential of the EU-Turkey readmission agreement and visa liberalization roadmap; upscaling the Poseidon Sea Joint Operation (Frontex, 2016) in Greece; reinforcing Frontex support at the borders between Bulgaria and Turkey, Greece-North Macedonia, Croatia-Serbia, deploying in Slovenia more police officers; strengthening the Frontex Western Balkan Risk Analysis Network (Frontex, 2020); making use, where appropriate, of the Rapid Border Intervention Team (RABIT) mechanism, that is a mechanism providing rapid operational assistance for a limited period to a requesting EU Member State facing a situation of urgent and exceptional pressure at points of the external EU borders from large numbers of third-country nationals trying to enter the territory of the EU Member State illegally.

The implementation of the EU-Turkey statement of March 2016 (European Council, 2016) has played a key role: in 2019 the arrivals via the Eastern Mediterranean route were 90 percent lower than in 2015, and a further decline was observed in 2020. At the same time, Turkey agreed to accept the rapid return of all migrants not in need of international protection crossing from Turkey to Greece and to take back all irregular migrants intercepted in the Turkish waters.

To date, the EU and Turkey continue this cooperation.

On 4 March 2020, the EU institutions expressed their solidarity with Greece, Bulgaria, Cyprus and other Member States similarly interested in efforts to manage the EU's external borders and support for the management of migration flows.

The EU and its Member States remain determined to protect the EU's external borders. Illegal crossing is not tolerated. They agree to take all necessary measures.

All Member States, the European Commission and EU Agencies stand ready to strengthen their support to areas under pressure through the deployment of Frontex rapid border intervention and additional technical assistance.

2.1 The implications of European migration policies on migration flows along the Turkish-Greek border

To collect data on migration along Balkan route is an extremely arduous work because of the participation of several countries. Migrants are counted many times during their journeys or, on the contrary, are not identified. For this work, I have used the official data of different international organisations, such as UNHCR and the Greek Agency Aegean Boat Report. I chose to analyse the situation at the Turkish-Greek border because it is the main entry point to the European border.

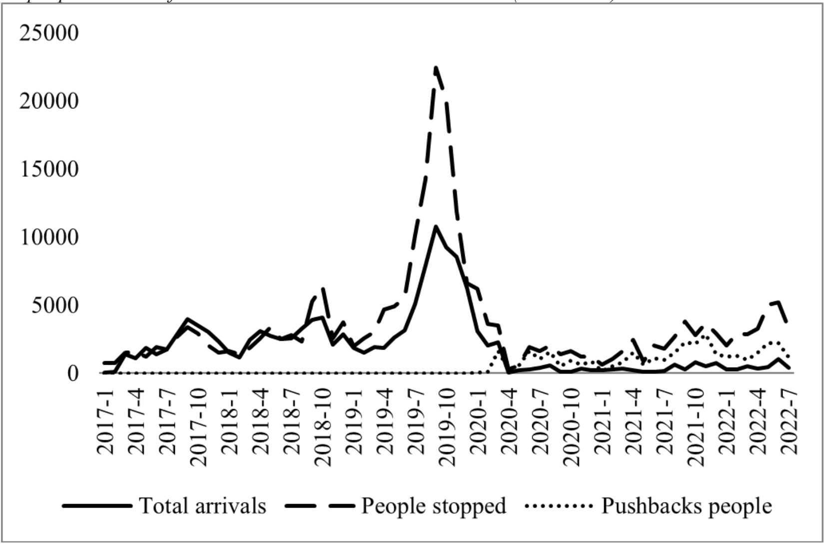
After a peak, for about a year, during which migrants at the European doors have used the formalized corridor to reach Europe, data show a drastic drop in arrivals (Figure 2). Thus, we can state that European policies effectively reflect the objectives, but what are their implications?

Including the last update on July 2022, the overall numbers registered since 2017 are:

- 133261 people arrived and 242102 people stopped;
- 130162 people transferred from the Greek Islands to mainland and a max population in camps on the islands equal to 42800.

According to the Aegean Boat Report (2022), in 2021 people arriving has decreased 54.9%, compared to 2020.

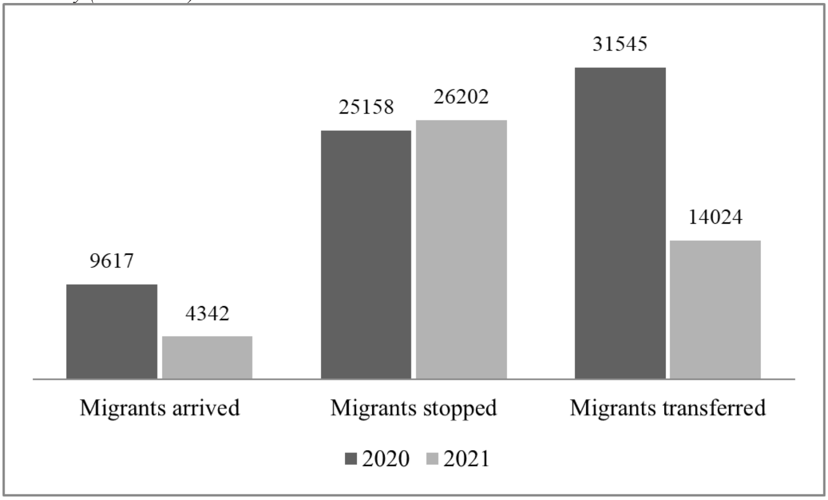
Figure 2. Trends of people arrived at the Greek Islands, people stopped by Turkish police and people pushed back from Greek Islands into Turkish waters (2017-2022)



Source: Aegean Boat Report Data Studio

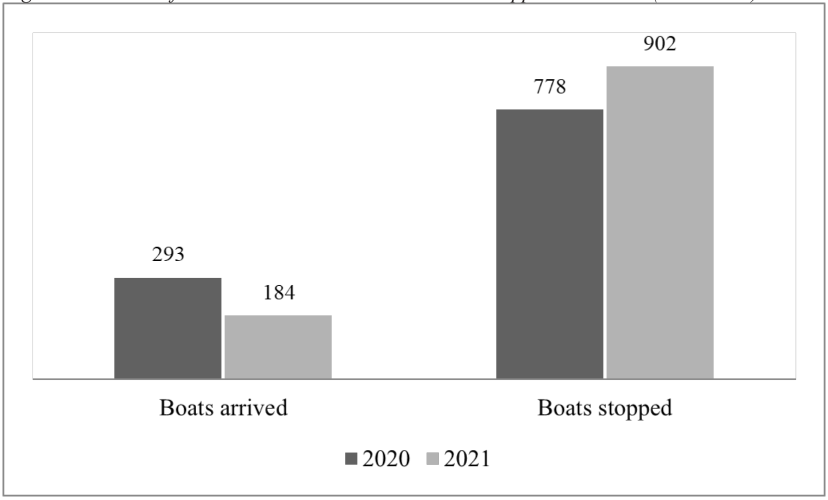
The study carried out by Aegean Boat association shows the trend of people arriving by boat and those stopping at the Greek Islands-Turkey border (Figure 3). Figure 4, on the other hand, refers to the number of boats that arrived at Greek Islands and stopped at borders during the same years (Aegean Boat Report, 2022).

Figure 3. Number of migrants arrived at the Greek Islands, stopped at borders and transferred to Turkey (2020-2021)



Source: Aegean Boat Report (2022)

Figure 4. Number of boats arrived at Greek Islands and stopped at borders (2020-2021)



Source: Aegean Boat Report (2022)

In 2020, of the 1071 boats that started their trip towards Greek Islands, only 293 arrived at their destination carrying 9617 people out of a total of 31545. In 2021, 184 boats made the trip, carrying 4342 people, the rest (902 boats with 26202 people) were picked up and arrested by the Turkish Coast guard and police. These data collected by Aegean Boat Report are in accordance with the number recorded by UNHCR equal to 9714 arrivals by sea in 2020 and 4331 arrivals by sea in 2021 (UNHCR, 2021b). Compared to 2020, it's easy to see how the arrivals have dropped but stopped boats and so stopped people have increased.

In 2021, 14024 people were transported to the mainland, from the Greek Islands, 55.5% less than in 2020. The total number of people hosted on Greek Islands has also decreased: official capacity is 13811 people. Today's camps population is less than in 2020 when it was 17005 and one year later it was equal to 3508, approximately 80% less. Camps population decreased by 13497 people last year by illegal pushbacks, violating international laws and human rights (Aegean Boat Report, 2022).

3 Conclusion

This article aims at identifying the role of Europe in managing migration policy along the Balkan countries. The study wants to outline the migrants' dangerous journeys across borders, partially legalized by European migration policies. The response to the migration crisis has been affecting Europe since 2015, although in line with a strong cooperation among countries, it fails in relation to its fundamental values and principles. From the analysis it emerges a series of violations of human rights informally admitted by the Member States, through the migration strategies established during the years. The externalization of borders and the creation of reception centres, in which refugees spend their time in a limbo situation, waiting for the possibility to apply for the asylum's request, represent the main wounds of Europe. The history of migration has a long genesis and demonstrates that it is impossible to stop the migration flows. The process of globalization has increased the migration flows around the world and the closure of borders cannot be the solution. On the contrary, the forced closure of one border coincides with a new harder path to reach the Western societies in which migrants hope to find security and better economic opportunities.

As demonstrated by data on arrivals, the absence of legal pathways in Europe for people in need of protection forces migrants to take other pathways, creating conditions even more dangerous.

The situation in the Balkan region reveals that the cooperation among States can reach its aim only when it is based on a solidarity vision, supported by migration policies of inclusion both within and outside the European countries. The EU-Turkey agreement is the first challenge that Europe must overcome in view of the number of migrants that continue to cross the borders, fuelling the smugglers. Camps, both formal and informal, should be temporary living spaces but they became permanent spaces of transit, structured by governments so to deny the basic rights of a civilised society. Increasingly, these realities so structured are used by government as political tools, which, heedless of human rights, repeatedly, trample the basis of a “civil” society.

Europe plays a fundamental role in these actions because of its financing of technical equipment and training used along the border. An important advancement that the Member States should foresee is the end of the externalization policy of borders by which both the border police of Balkan countries and Frontex, the European agency, pushback migrants from one country to another in violation of human rights.

References

Abikova J. and Wojciech P. (2021), Shaping the Balkan Corridor: Development and Changes in the Migration Route 2015–16, *International Migration,* 59(5), 248-65, available at https://doi.org/10.1111/imig.12828

AEGEAN BOAT REPORT, AEGEAN BOAT REPORT Data Studio, available at https://datastudio.google.com/reporting/1CiKR1_R7-1UbMHKhzZe_Ji_cvqF7xlfH/page/A5Q0 [4 May, 2022].

AEGEAN BOAT REPORT (2022), Annual Report 2021, available at https://aegeanboatreport.com/reports/ [4 May, 2022].

Ahmetašević N. (2020), Limits to Access to Asylum Along the Balkan Route, *Mešić S. Editor,* available at https://refugee-rights.eu/wp-content/uploads/2020/07/RRE_LimitedAccessToAsylumAlongTheBalkanRoute.pdf

El-Shaarawi N. and Razsa M. (2018), Movements upon Movements: Refugee and Activist Struggles to Open the Balkan Route to Europe, *History and Anthropology* 30(1), 91-112, available at https://doi.org/10.1080/02757206.2018.1530668

EUROPEAN COMMISSION (2015a), A European Agenda on Migration, 13 May, available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52015DC0240&from=EN [23 March, 2022].

EUROPEAN COMMISSION (2015b), Meeting on the Western Balkans Migration Route: Leaders Agree on 17-Point Plan of Action, Press Release, 25 October, available at https://ec.europa.eu/commission/presscorner/detail/en/IP_15_5904 [6 April, 2022].

EUROPEAN COMMISSION (2016), EU-Turkey Joint Action Plan: Implementation Report, 10 February, available at https://ec.europa.eu/homeaffairs/system/files/2016-12/managing_the_refugee_crisis_-_eu-turkey_join_action_plan_implementation_report_20160210_en.pdf [6 April, 2022]

EUROPEAN COUNCIL (2016), EU-Turkey Statement, Press Release, 18 March, available at https://www.consilium.europa.eu/en/press/press-releases/2016/03/18/eu-turkey-statement/, [6 April, 2022].

FRONTEX (2016), Joint Operation Poseidon (Greece), Focus, 10 October, available at https://frontex.europa.eu/we-support/main-operations/operation-poseidon-greece-/ [6 April, 2022].

FRONTEX (2020), Situational awareness and monitoring, Strategic Analysis, available at https://frontex.europa.eu/we-know/situational-awareness-and-monitoring/strategic-analysis/ [6 April, 2022].

Hameršak M., Hess S., Speer M., Stojić Mitrović M. (2020), The Forging of the Balkan Route. Contextualizing the Border Regime in the EU Periphery, *movements: Journal for Critical Migration and Border Regime Studies*, 5(1), 9-19, available at https://movements-journal.org/issues/08.balkanroute/01.hamersak,hess,stojic-mitrovic,speer--the-forging-of-the-balkan-route.html

Ibrahim Y. (2018), The Unsacred and the Spectacularized: Alan Kurdi and the Migrant Body, *Social Media + Society,* 4(4), 1-9, available at https://doi.org/10.1177/2056305118803884

MEDECINS SANS FRONTIERES (2018), *Balkans: International Activity Report 2018,* available at https://www.msf.org/international-activity-report-2018/balkans.

MEDECINS SANS FRONTIERES (2022), MSF Leaves Polish Border after Being Blocked from Assisting People, Press Release, 6 January, available at https://www.msf.org/msf-leaves-polish-border-after-being-blocked-assisting-migrants-and-refugees [4 May, 2022].

Sapoch J. and Augustova K., (2020), Border Violence as Border Deterrence Condensed Analysis of Violent Push-Backs from the Ground, *Politics, History and International Relations*, 5(1), 1-20, available at https://movements-journal.org/issues/08.balkanroute/12.augustova,sapoch--border-violence-as-border-deterrence.pdf

Stojić Mitrović M., Speer M. and Beznec B. (2016), Governing the Balkan Route: Macedonia, Serbia and the European Balkan Regime, Rosa Luxemburg Stiftung Southeast Europe, available at https://www.rosalux.de/en/publication/id/14554/governing-the-balkan-route

UNHCR (2021a), UNHCR OPERATIONAL DATA PORTAL, Mediterranean Situation, available at https://data.unhcr.org/en/situations/mediterranean/location/5179 [4 May, 2022].

UNHCR (2021b), UNHCR OPERATIONAL DATA PORTAL, Greece Sea Arrivals Dashboard, December 2021, available at https://data2.unhcr.org/en/documents/details/90834 [4 May, 2022].

UNHCR OPERATIONAL DATA PORTAL (2022), Western Balkans - Refugees, asylum-seekers and other people in mixed movements, available at https://data.unhcr.org/en/documents/details/96520 [4 May, 2022].

Acknowledgements

This work represents the crucial part of a broader research project started a year ago for the writing of my Master's thesis. It would not have been possible without the help of my supervisor, Professor Marina Zannella. I thank her not only for guiding me in my research but above all for stimulating me during her course at Sapienza University of Rome.

I also thank Professor Alessandra De Rose, for her invaluable help with suggestions and revisions of the text.

Finally, I would like to thank all the people who, through their stories, have inspired me to study and delve into the dynamics of society.

SAPIENZA
UNIVERSITÀ EDITRICE

ANNALI DEL DIPARTIMENTO DI METODI E MODELLI
PER L'ECONOMIA IL TERRITORIO E LA FINANZA
2022 | pp. 151-156

E-ISSN 2611-6634 | ISSN 2385-0825
DOI: 10.13133/2611-6634/1440

Notes and discussions

Marina Zannella[*]

RESIDENZIALITÀ INCLUSIVA E SILVER CO-HOUSING: L'UNIONE FA LA FORZA

Abstract

Long-term demographic dynamics, characterised by increasing life expectancy, low fertility and marital instability, pose major challenges to the Italian welfare, particularly with regard to the care of the elderly and their well-being. Ageing in place and inclusive housing are key elements to foster and promote active ageing. However, in Italy solidarity and silver cohousing projects are still scarce and limited to the sphere of social experimentation. This article presents existing good practices on the territory, which could represent a starting point for rede-signing the Italian housing system in the direction of greater inclusiveness.

Keywords: active ageing, ageing in place, welfare, cohousing.

---

[*] Dipartimento Metodi e Modelli per il Territorio, l'Economia e la Finanza (MEMOTEF) - Sapienza Università di Roma

1 L’inverno demografico italiano

Secondo gli ultimi dati diffusi dall'Istat per il 2019 (Istat, 2020), in Italia ci sono oltre 7 milioni di persone con più di 75 anni, pari all'11,7 per cento della popolazione, mentre più di 4 milioni hanno spento almeno 80 candeline. L'invecchiamento della popolazione è un fenomeno destinato a intensificarsi in futuro ed entro il 2050 gli anziani con più di 65 anni potrebbero rappresentare il 35% della popolazione italiana, rendendo ancora più necessario un adeguamento delle politiche sociali.

Le dinamiche demografiche in atto porteranno a importanti cambiamenti anche nelle strutture familiari, uno fra tutti l'aumento delle micro-famiglie unipersonali. Attualmente, gli anziani soli rappresentano la metà di tutte le famiglie composte da una sola persona e i dati Istat più recenti stimano che, sul totale degli anziani, circa uno su tre vive da solo.

Nonostante la tendenza in aumento delle persone sole, l'Italia continua ad essere caratterizzata dall’esistenza di forti legami intergenerazionali: il 20,9% degli anziani over 75 vive con i figli, mentre tra coloro che hanno figli e vivono da soli, il 56,9% è abituato a incontrarli quotidianamente. Tra i genitori anziani che vivono da soli, il 79,4% abita nello stesso comune dei figli, il 15,1% vive nello stesso condominio e il 25,8% abita entro 1 km di distanza.

La vicinanza abitativa e il contatto tra genitori e figli sono fenomeni più diffusi nei Paesi dell'Europa meridionale, caratterizza-ti da legami familiari più stretti, rispetto agli altri Paesi europei. Queste differenze possono essere almeno in parte attribuite ai diversi sistemi di welfare, in particolare per quanto riguarda l'organizzazione del lavoro di cura e del sostegno intergenerazionale, una funzione ampiamente delegata dallo Stato alle famiglie in Italia e in altri Paesi mediterranei (Ferrera, 1996). Ma cosa accadrà in futuro?

Le dinamiche demografiche di lungo periodo legate all'aumento dell'aspettativa di vita, alla bassa fecondità e all'instabilità coniugale incideranno sulla capacità delle famiglie di fornire sostegno intergenerazionale a causa della riduzione del numero medio di componenti (con un aumento delle coppie senza figli a scapito di quelle con figli) e dell'aumento delle famiglie senza nucleo, caratterizzate dall'assenza di relazioni di parentela di coppia o del tipo figlio-genitore.

L’inverno demografico italiano rende, pertanto, necessario ripensare i modelli di assistenza che dovranno fare i conti con un numero crescente di anziani a fronte di famiglie sempre più ridotte e frammentate e con la progressiva flessione delle giovani gene-razioni. Allo stesso tempo, il progressivo invecchiamento della popolazione e l'aumento della domanda di assistenza dovuta all'insorgere di patologie e alla perdita di autonomia in età avanzata rappresentano una sfida non solo in termini di offerta sanitaria ma anche per quanto riguarda le politiche abitative.

L’indagine europea sulla salute (*European Health Interview Survey*, EHIS, Eurostat, 2019a) traccia un quadro preoccupante dell’attuale domanda di assistenza in Italia che, secondo il Rapporto della Commissione per la riforma dell’assistenza sanitaria e sociosanitaria per la popolazione anziani (Istat, 2021), tra gli individui dai 75 anni in su, assume un peso rilevante a causa delle ridotte capacità funzionali, della carenza di supporto sociale, della necessità di sostegno, delle condizioni abitative sfavorevoli e delle difficoltà economiche. I dati EHIS mostrano infatti che, su 6,9 milioni di over 75, oltre 2,7 milioni presentano gravi difficoltà motorie, comorbilità e limitazioni dell’autonomia nello svolgimento di attività necessarie per la vita quotidiana, quasi la metà (1,2 milioni di anziani) dichiara di non poter contare su un aiuto adeguato alle proprie necessità, e tra questi circa 1 milione vive solo oppure con altri familiari anziani. Le statistiche sull'offerta di assistenza a lungo termine dei pazienti anziani cronici (*Long-Term Care*) evidenziano, inoltre, l'arretratezza dell'Italia che si colloca tra gli ultimi posti della classifica dei Paesi OCSE (OECD, 2019) per disponibilità di posti letto in strutture residenziali di *long-term care*. I posti disponibili in Italia al 2019 sono infatti 18,8 ogni mille individui con più di 65 anni, un valore molto basso se confrontato con altri Paesi del Nord Europa (80,8 in Lussemburgo, 72,1 nei Paesi Bassi e 68,7 in Belgio) ma anche con altri Paesi mediterranei (43,9 in Spagna).

Questa situazione potrebbe essere aggravata dalla condizione economica della popolazione anziana. Secondo il "Rapporto Abitare Anziani" (Falasca, 2015), l'80,3% degli anziani vive in case di proprietà; spesso, tuttavia, le condizioni abitative sono inadeguate a causa delle difficoltà economiche a intraprendere interventi di ristrutturazione e adeguamento. Dal rapporto si evince che nel 70% dei casi si tratta di abitazioni di almeno 50 anni; il 7% di queste non sono dotate di impianto di riscaldamento e nel 56% dei casi sono a un piano superiore al secondo ma non dotato di ascensore. Inoltre, uno sguardo alle statistiche Eurostat rivela che, nel 2019, in Italia, il 27,5% degli over 65 che vivono da soli non può permettersi di sostituire i mobili usurati, un dato superiore alla media europea (26,6%) e lontano da quello registrato in molti Paesi dell'Europa centrale e settentrionale come, ad esempio, la Svezia con una percentuale del 4,9% (Eurostat, 2019b).

Alla luce del quadro complessivo finora descritto, si comprende come sia fondamentale ai fini del benessere individuale e sociale identificare e soddisfare i bisogni degli anziani, specialmente di quelli più vulnerabili, attraverso un generale miglioramento delle condizioni di vita che passa, in primo luogo, dalla possibilità di vivere in alloggi idonei a garantire quanto più possibile una vita autonoma, di ricevere adeguati livelli di assistenza e cura e di poter partecipare attivamente alla vita sociale. Questo articolo intende offrire una panoramica sulle origini e sulla diffusione delle pratiche residenziali inclusive rivolte agli anziani, con l'obiettivo principale di tracciare lo stato dell'arte in Italia, evidenziando buone pratiche e aspetti critici.

2 Invecchiamento attivo e residenzialità inclusiva

Il Piano d'azione internazionale sull'invecchiamento di Madrid (MIPPA), adottato nel 2002, include tra i suoi obiettivi la necessità di promuovere i processi di *aging in place* e *aging in community*, che garantiscano opzioni abitative accessibili per le persone anziane tenendo conto delle preferenze individuali (United Nations, 2002). Tra le azioni prioritarie da intraprendere per raggiungere questo obiettivo, il MIPPA raccomanda la promozione di varie forme di *cohousing* (ad esempio: inter e intra-generazionale, *cohousing* di quartiere, condomini solidali e villaggi eco-rurali, *housing* sociale, ecc.) in età anziana e forme più innovative di rigenerazione urbana, per favorire sistemi di convivenza sociale in grado di stimolare una partecipazione attiva e di promuovere una vita indipendente in considerazione delle esigenze delle persone anziane, in particolare di quelle con disabilità.

Il *cohousing* nasce in Danimarca alla fine degli anni Sessanta ad opera dell'architetto Jan Gødmand Høyer, come nuovo modo di vivere insieme abitando in alloggi privati ma condividendo spazi e tempo, in risposta al disagio abitativo e alle difficoltà nelle relazioni sociali osservati nei grandi condomini di Copenaghen. Negli anni Settanta il *cohousing* si è diffuso anche in altri paesi, tra cui Svezia, Olanda, Regno Unito e Sati

Uniti, con il principale scopo di migliorare i livelli di benessere individuale, contrastando la rarefazione e l'inaridimento delle relazioni, ma anche il benessere sociale attraverso la riduzione degli sprechi ambientali, il risparmio e il rafforzamento del senso di comunità (Cummings and Kropf, 2020).

In anni più recenti, a seguito del processo di invecchiamento in atto nella popolazione dei paesi occidentali, si è registrato un crescente interesse per le soluzioni di *cohousing* rivolte agli anziani, talvolta chiamate *senior cohousing*, *silver* (*co*)*housing* o *senior living*. Nei paesi europei dove l'investimento nel *senior cohousing* è maggiore (ad esempio Regno Unito, Germania e Francia), sono emerse diverse esperienze, tra cui: alloggi in aree residenziali per anziani, residenze per anziani con servizi di assistenza che possono includere una casa di riposo al loro interno, condomini solidali. Una caratteristica comune alla maggior parte di questi progetti è quella di mirare a promuovere il senso di comunità locale e favorire la solidarietà intergenerazionale per contrastare l'isolamento e la solitudine in età avanzata. Ne è un esempio, per certi versi radicale, la comunità di Sällbo[1] nella città di Helsingborg, in Svezia, che ha trasformato una ex casa di riposo in un vivace complesso abitativo in cui la metà dei residenti ha più di 70 anni e il resto ha un'età compresa tra 18 e 25 anni. I residenti sono stati individuati dopo un accurato processo di selezione mediante intervista per garantire un mix di personalità, background, religioni e valori, e tutti hanno dovuto firmare un contratto in cui si impegnavano a trascorrere almeno due ore alla settimana a socializzare con i loro vicini. Un altro esempio è il progetto di *senior cohousing* denominato *Biloba*, nella città di Bruxelles, in cui oltre ad alloggi privati per anziani sono previste anche strutture comuni aperte a tutto il quartiere al fine di favorire la creazione ed il raf-forzamento delle reti sociali.

3 Esperienze di *cohousing* solidale e *silver cohousing* in Italia

Non esiste ad oggi una banca dati per l'Italia sul numero di persone che risiedono in *silver cohousing* né in altre forme di *housing* sociale. Le informazioni al riguardo, di natura frammentaria, sono spesso difficili da reperire. Recentemente, lo sviluppo di sperimentazioni di forme residenziali inclusive è stato favorito dall'entrata in scena nel settore dell'edilizia residenziale sociale di nuovi operatori, principalmente fondazioni per lo sviluppo del *social housing*, oltre al settore pubblico e alle cooperative edilizie private.

Da una recente rassegna condotta all'interno delle amministrazioni regionali nell'ambito del progetto di "Coordinamento nazionale partecipato e multilivello delle politiche sull'invecchiamento attivo" (Lucantoni e Principi, 2022), promosso dal Dipartimento per le Politiche della Famiglia, indica, la Regione Piemonte -e, in particolare, il Comune di Torino- tra i principali forieri di *best practice* in ambito di città sostenibili e *housing* sociale. Tali progetti innovativi, nati dalla collaborazione tra soggetti pubblici e privati, mirano a migliorare la qualità della vita degli individui sviluppando

---

[1] Informazioni sulle *best practice* in materia di *cohousing* e *silver cohousing* presenti in Europa possono essere trovate alle seguenti pagine web: https://www.age-platform.eu/; https://www.housingeurope.eu/.

risposte diversificate per le esigenze e la necessità delle diverse fasi del ciclo di vita. I progetti, rivolti anche alla popolazione anziana, sono diversi e includono il *cohousing* tra diverse tipologie di famiglie (persone singole, coppie, coppie con figli, etc.) che vivono in appartamenti privati ma condividono spazi comuni destinati sia allo svolgimento di attività gestionali (cucina, lavanderia, sala riunioni) che alle attività del tempo libero (laboratori per il fai da te, sale per il gioco)[2]. I *cohousers* non si limitano a condividere spazi ma svolgono a turno servizi utili per la comunità ed il vicinato, tra cui la sorveglianza dei bambini, la spesa settimanale, la cura del verde e la manutenzione ordinaria degli edifici.

Tra le sperimentazioni del Comune di Torino ci sono anche progetti di *cohousing solidale*, ovvero forme di coabitazione intergenerazionale in cui i giovani, in cambio di una riduzione delle spese d'affitto, prestano ore di volontariato per migliorare le relazioni tra gli inquilini dello stabile, offrono sostegno e accompagnamento alle persone più fragili e collaborano alla piccola manutenzione delle parti comuni. Un importante esempio di coabitazione solidale è *To Housing*, il primo progetto di *cohousing* sociale rivolto all'accoglienza di persone LGBT in grave difficoltà. Il progetto è destinato a giovani tra i 18 e 26 anni allontanati dalle famiglie di origine a causa dell'orientamento sessuale, anziani LGBT in condizione di solitudine o povertà, migranti e rifugiati omosessuali. Un'esperienza analoga alla coabitazione solidale attiva sul territorio è quella dei condomini solidali destinati a gruppi di popolazione vulnerabile. Si tratta di un modello di convivenza basato sulla solidarietà intergenerazionale che prevede, oltre all'offerta di alloggi privati e spazi condivisi, anche forme di mutuo sostegno talvolta con il supporto di volontari e operatori professionali.

Se il *cohousing sociale* è ancora poco diffuso in Italia rispetto ad altri Paesi europei, le esperienze di *silver cohousing* sono ancora più limitate[3]. Tra queste il *Cohousing del Moro*, sito nel centro della città di Lucca, è destinato a persone anziane autosufficienti e prevede soluzioni abitative diversificate, spazi condivisi per la socialità, un'offerta di servizi che varia dall'attività di primo soccorso alla presenza di ambulatori medici alla promozione di attività culturali. Nel 2015 è stata, inoltre, istituita a Roma la Residenza "Giuseppe e Margherita Achillini", dotata di sei appartamenti autonomi con servizi collettivi destinati a coppie di anziani che si trovano in una situazione di imminente comprovato rischio di espulsione dal proprio contesto abitativo e che non siano in grado di trovare una soluzione abitativa alternativa per documentate ragioni economiche. L'obiettivo è quello di evitare lo sradicamento delle persone anziane dal proprio quartiere ed ambiente di vita a causa dei processi di espulsione connessi all'aumento dei canoni di locazione, alle azioni di sfratto e/o ai processi di alienazione del patrimonio immobiliare pubblico, assicurativo, previdenziale e offrire una risposta di tipo comunitario diversa dalla istituzionalizzazione, garantendo il mantenimento della propria autonomia. Altri due esempi di *silver cohousing* sono *Casa Tassullo* e *Casa Cles*, in provincia di Trento, nate a seguito dell'esperienza di *cohousing* intergenerazionale *Casa alla Vela*, una delle più note convivenze solidali tra giovani e anziani in Italia, istituita all'interno dello stesso territorio nel 2014.

---

[2] Una descrizione dei progetti presenti nel Comune di Torino è presente online: http://www.comune.torino.it/torinogiovani/vivere-a-torino/housing-sociale.

[3] Per una rassegna dei progetti di *cohousing* e *silver cohousing* presenti in Italia si consulti anche: https://ioabitosocial.it/la-piattaforma-di-housing-sociale-in-italia.

4 Riflessioni conclusive

In Italia, come in altri Paesi europei, la pandemia ha messo a nudo la fragilità dei sistemi di assistenza agli anziani incentrati sull'offerta di presidi residenziali socio-assistenziali e socio-sanitari. L'importanza di questa tematica per lo sviluppo sostenibile del Paese è messa in evidenza anche all'interno del Piano Nazionale di Ripresa e Resilienza (PNNR) che prevede, nella Missione 5 Inclusione e coesione, due fonti di finanziamento che riguardano le condizioni abitative di anziani e disabili con l'obiettivo di prevenirne l'istituzionalizzazione anche attraverso l'offerta di alloggi. Tuttavia, come evidenziato dalla Caritas (2022), gli interventi sul piano abitativo previsti dal PNRR rischiano di non intercettare pienamente l'obiettivo dell'*ageing in place* a causa dell'assenza di una *governance* e di un piano condiviso per l'integrazione delle diverse linee e forme di finanziamento (nazionali, regionali e locali) necessari per ridisegnare il sistema di *housing* italiano, storicamente "poco sociale", nella direzione di una maggiore inclusività.

References

CARITAS (2022), Casa e abitare nel PNNR. Analisi e prospettive. *Quaderni sulla ripresa e resilienza del Paese*, n. 1, marzo 2022.

EUROSTAT (2019a), *European Health Interview Survey – EHIS 2019.*

EUROSTAT (2019b), *Statistics on Income and Living Conditions.*

Falasca C. (a cura di) (2015), *Secondo Rapporto sulle condizioni abitative degli anziani in Italia che vivono in case di proprietà*. Auser Nazionale, Abitare e anziani, Roma: GGIL_SPI Roma.

Ferrera M. (1996), The Southern model of welfare in social Europe, *Journal of European social policy*, *6*(1), 17-37.

ISTAT (2020), Aspetti della vita degli over 75, disponibile online: https://www.istat.it/it/files/2020/04/statisticatoday_ANZIANI.pdf

ISTAT (2021), Gli anziani e la loro domanda sociale e sanitaria. Anno 2019. Rapporto Commissione per la Riforma dell'assistenza sanitaria e sociosanitaria per la popolazione anziana. https://www.istat.it/it/files//2021/06/rapporto_commissione_anziani.pdf

Cummings S. and Kropf N. P. (2020), *Senior Cohousing—History and Theory*, In *Senior Cohousing* (pp. 9-16). Springer, Cham.

Lucantoni D., Principi A. (a cura di) (2022), *Politiche per l'invecchiamento attivo in Italia: Quali possibili obiettivi nelle Regioni, nelle Province Autonome, nei Ministeri e nei Dipartimenti presso la Presidenza del Consiglio dei Ministri? Raccolta di rapporti.* Progetto "Coordinamento nazionale partecipato e multilivello delle politiche sull'invecchiamento attivo", Work Package 2 – Task 3. Disponibile online: https://famiglia.governo.it/media/2704/possibili-obiettivi-politiche-per-linvecchiamento-attivo-in-italia-raccolta-dei-rapporti.pdf

OECD (2019), Data on Long-Term Care Resources and Utilisation

UNITED NATIONS (2002), Political Declaration and Madrid International Plan of Action on Ageing; United Nations: New York, NY, USA, 2002. Disponibile online: https://www.un.org/esa/socdev/documents/ageing/MIPAA/political-declaration-en.pdf

BOOK REVIEWS

SAPIENZA
UNIVERSITÀ EDITRICE
ANNALI DEL DIPARTIMENTO DI METODI E MODELLI
PER L'ECONOMIA IL TERRITORIO E LA FINANZA
2022 | pp. 159-160

E-ISSN 2611-6634 | ISSN 2385-0825
DOI: 10.13133/2611-6634/1400

Book Review

Alessandra De Rose*

MINELLO ALESSANDRA, NON È UN PAESE PER MADRI, EDITORI LATERZA, 2022, PP. 150

L'Italia è entrata nel suo inverno demografico: nascono sempre meno bambini mentre il numero di morti è in continuo aumento per effetto dell'invecchiamento strutturale della popolazione e le immigrazioni dall'estero di cittadini stranieri non sono sufficientemente numerose da compensare il saldo negativo tra natalità e mortalità. Il risultato è una popolazione in declino, destinata a diminuire ancora da qui al 2060 – financo a dimezzare, nelle più pessimistiche previsioni – con conseguenze gravi sulla vitalità economica e sociale del nostro paese. I demografi da tempo hanno segnalato il pericoloso andamento delle dinamiche della popolazione italiana, ne hanno esplorato a fondo le cause, prefigurato le conseguenze e indicato le possibili contromisure, che, attraverso azioni di policy illuminata e duratura, potrebbero rallentare la corsa verso il declino demografico se non di invertire la rotta, così come accaduto in altri paesi europei.

Gi obiettivi di questo denso, ma agile volume di Alessandra Minello sono proprio questi: da un lato, illustrare i dati relativi al fenomeno del declino delle nascite che è all'origine della crisi della popolazione italiana e del suo invecchiamento strutturale, approfondendo tutti gli aspetti che rendono difficile se non impossibile la maternità nel nostro Paese; dall'altro, riflettere molto pragmaticamente sulle soluzioni possibili, sulle azioni positive da mettere in campo affinché l'Italia possa tornare ad essere "un paese per madri".

Le argomentazioni partono dalla constatazione del *fertility gap* cioè dalla differenza tra numero medio di figli desiderato e numero medio di nascite effettivamente osservate. Mentre il primo indicatore si mantiene stabilmente sopra il valore di 2, il secondo, ovvero la fecondità realizzata, è in continua diminuzione da decenni, oggi intorno a 1,3 figli per donna. Poiché, inoltre, la percentuale di donne che dichiarano di voler rimanere senza figli è molto bassa mentre cresce quella che hanno figli ad età elevate, oltre i 40 anni – e, di conseguenza, quella di coloro che si rivolgono alle tecniche di fecondazione

---

* Sapienza University of Rome

medicalmente assistita - è evidente che nel nostro paese c'è una larga quota di maternità insoddisfatta.

Per spiegare perché una dimensione della vita, che riveste ancora un ruolo centrale per la maggior parte delle donne è diventata sempre più complicata, occorre richiamare sia aspetti culturali che aspetti strutturali, organizzativi, economici, che pesano sulle spalle delle italiane.

Tra i primi, il *mito della maternità* che, secondo l'Autrice, sostenendo il desiderio di avere figli al di là della spinta biologica, ammanterebbe eccessivamente le donne del senso di responsabilità già a partire dalla gestione della gravidanza e del parto, ma soprattutto nella fase della cura e dell'allevamento dei figli che, nell'immaginario collettivo, richiede presenza ed abnegazione soprattutto da parte della madre. Secondo questo approccio, le donne, spinte dalla paura di non essere all'altezza del ruolo – che nel frattempo è entrato in competizione con altri ruoli, primo fra tutti quello lavorativo – rimanderebbero la scelta riproduttiva fino a rinunciarvi del tutto. Se la donna è lasciata sola di fronte a queste scelte e all'impegno che ne deriva, il rischio che rimanga senza figli o che ne abbia al più uno solo è molto alto. Ecco perché un'altra questione culturale che va affrontata è quella del coinvolgimento del partner nella cura dei figli e nel lavoro familiare. Nel nostro paese, come la Minello spiega molto bene, la *gender revolution*, cioè il processo di riequilibrio del potere tra uomini e donne non è ancora completa. Tale processo si compone, infatti, di due fasi: nella prima è la donna a conquistare un posto nello spazio pubblico, partecipando al mercato del lavoro e ricoprendo ruoli sempre più importanti nei luoghi di potere. Nella seconda fase è l'uomo l'agente attivo, che comincia ad assumere un ruolo sempre più partecipe alla vita domestica, condividendo con la donna compiti e decisioni nel lavoro di cura. In Italia siamo fermi alla prima fase, sulla quale peraltro "*arranchiamo*", non solo per un'arretratezza culturale, ma anche, e forse soprattutto, per carenze strutturali di fondo, che rappresentano il secondo importante ordine di spiegazioni della bassa fecondità evocate dall'Autrice.

Il difficile rapporto tra maternità e lavoro riassume gran parte di queste difficoltà. Ad esso è dedicato un intero capitolo, ma, inevitabilmente, il tema è richiamato in tutto il libro: l'ingresso delle donne italiane nel mercato del lavoro, iniziato in ritardo rispetto ad altri paesi europei e non ancora completato, non è stato accompagnato da misure adeguate di sostegno alla maternità – nonostante una delle leggi più generose per quanto riguarda la tutela delle lavoratrici madri – perché i servizi per l'infanzia non sono della numerosità e della qualità attesa, perché l'organizzazione del lavoro è rigida e la flessibilità offerta alle donne si segnala più come portatrice di precarietà che come una soluzione accettabile, perché ci si deve affidare ancora troppo alla solidarietà intergenerazionale privata, perché non si è fatto ancora abbastanza per completare la *gender revolution*.

Tutti questi aspetti sono affrontati nel libro con ricchezza di dati e di esempi e anche di proposte concrete, che le istituzioni dovrebbero mettere in atto per aiutare le donne a superare le criticità "*…immaginando una società in cui vita professionale e vita privata siano in armonia*". Può sembrare uno slogan femminista, così come tanti dei percorsi indicati sono – a detta della stessa Autrice – ideali da raggiungere, tortuosi e con tanti ostacoli da superare. Tuttavia, una società che non assicura il soddisfacimento dei bisogni individuali e, in particolare, il desiderio di maternità – e di paternità - è una società che ha fallito il suo mandato e che è destinata al declino.

ISSN: 2385-0825 (print)
ISSN: 2611-6634 (online)

www.editricesapienza.it