# Chapter 2
# Uncertainty and Complexity: Towards Model-Based Demography

**Jakub Bijak**

This chapter focuses on the broad methodological and philosophical underpinnings of the Bayesian model-based approach to studying migration. Starting from reflections on the uncertainty and complexity in demography and, in particular, migration studies, the focus moves to the shifting role of formal modelling, from merely describing, to predicting and explaining population processes. Of particular importance are the gaps in understanding asylum migration flows, which are some of the least predictable while at the same time most consequential forms of human mobility. The well-recognised theoretical void of demography as a discipline does not help, especially given the lack of empirical micro-foundations in formal modelling. Here, we analyse possible solutions to theoretical shortcomings of demography and migration studies from the point of view of the philosophy of science, looking at the inductive, deductive and abductive approaches to scientific reasoning. In that spirit, the final section introduces and extends a research programme of model-based demography.

## 2.1 Uncertainty and Complexity in Demography and Migration

The past, present, and especially the future size and composition of human populations are all, to some extent, uncertain. Population dynamics results from the interplay between the three main components of population change – mortality, fertility and migration – which differ with regard to their predictability. Long-term trends indicate that mortality is typically the most stable and hence the most predictable of the three demographic components. At the same time, the uncertainty of migration is the highest, and exhibits the most volatility in the short term (NRC, 2000).

Next to being uncertain, demographic processes are also complex in that they result from a range of interacting biological and social drivers and factors, acting in

non-linear ways, with human agency – and free will – exercised by the different actors involved. There are clear links between uncertainty and complexity: for mortality, the biological component is very high; contemporary fertility is a result of a mix of biological and social factors as well as individual choice; whereas migration – unlike mortality or fertility – is a process with hardly any biological input, in which human choice plays a pivotal role. This is one of the main reasons why human migration belongs to the most uncertain and volatile demographic processes, being as it is a very complex social phenomenon, with a multitude of underpinning factors and drivers.

On the whole, uncertainty in migration studies is pervasive (Bijak & Czaika, 2020). Migration is a complex demographic and social process that is not only difficult to conceptualise and to measure (King, 2002; Poulain et al., 2006), but also – even more – to explain (Arango, 2000), predict (Bijak, 2010), and control (Castles, 2004). Even at the conceptual level, migration does not have a single definition, and its conceptual challenges are further exacerbated by the very imprecise instruments, such as surveys or registers, which are used to measure it.

Historically, attempts to formalise the analysis of migration have been proposed since at least the seminal work of Ravenstein (1885). Contemporarily, a variety of alternative approaches co-exist, largely being compartmentalised along disciplinary boundaries: from neo-classical micro-economics, to sociological observations on networks and institutions (for a review, see Massey et al., 1993), or macro-level geographical studies of gravity (Cohen et al., 2008), to 'mobility transition' (Zelinsky, 1971) and unifying theories such as migration systems (Mabogunje, 1970; Kritz et al., 1992), or Massey's (2002) less-known synthesising attempt.

At the same time, the very notions of risk and uncertainty, as well as possible ways of managing them, are central to contemporary academic debates on migration (e.g. Williams & Baláž, 2011). Some theories, such as the *new economics of migration* (Stark & Bloom, 1985; Stark, 1991) even point to migration as an active strategy of risk management on the part of the decision-making unit, which in this case is a household rather than an individual. Similar arguments have been given in the context of environment-related migration, where mobility is perceived as one of the possible strategies for adapting to the changing environmental circumstances in the face of the unknown (Foresight, 2011).

Still, there is general agreement that none of the existing explanations offered for migration processes are fully satisfactory, and theoretical fragmentation is at least partially to blame (Arango, 2000). Similarly, given meagre successes of predictive migration models (Bijak et al., 2019), the contemporary consensus is that the best that can be achieved with available methods and data is a coherent, well-calibrated description of uncertainty, rather than the reduction of this uncertainty through additional knowledge (Bijak, 2010; Azose & Raftery, 2015). Due to ambiguities in migration concepts and definitions, imprecise measurement, too simplistic attempts at explanation, as well as inherently uncertain prediction, it appears that the demographic studies of migration, especially looking at macro-level or micro-level processes alone, have reached fundamental epistemological limits.

Recently, Willekens (2018) reviewed the factors behind the uncertainty of migration predictions, including the poor state of migration data and theories, additionally pointing to the existence of many motives for migration, difficulty in delineating migration versus other types of mobility, and the presence of many actors, whose interactions shape migration processes. In addition, the intricacies of the legal, political and security dimensions make international migration processes even more complex from an analytical point of view.

The existing knowledge gaps in migration research can be partially filled by explicitly and causally modelling the individuals (agents) and their decision-making processes in computer simulations (Klabunde & Willekens, 2016; Willekens, 2018). In particular, as advocated by Gray et al. (2016), the psychological aspects of human decisions can be based on data from cognitive experiments similar to those carried out in behavioural economics (Ariely, 2008). Some of the currently missing information can be also supplemented by collecting dedicated data on various facets of migration processes. Given their vast uncertainty, this could be especially important in the context of asylum migration flows, as discussed later in this chapter.

## 2.2   High Uncertainty and Impact: Why Model Asylum Migration?

Among the different types of migration, those related to various forms of involuntary mobility, violence-induced migration, including asylum and refugee movements, have the highest uncertainty and the highest potential impact on both the origin and destination societies (see, e.g. Bijak et al., 2019). Such flows are some of the most volatile and therefore the least predictable. They are often a rapid response to very unstable and powerful drivers, notably including armed conflict or environmental disasters, which lead people to leave their homes in a very short period (Foresight, 2011). Despite the involuntary origins, different types of forced mobility, including asylum migration, like all migration flows, also prominently feature human agency at their core: this is well known both from scholarly literature (Castles, 2004), as well as from journalistic accounts of migrant journeys (Kingsley, 2016).

As a result, and also because it is difficult to disentangle asylum migration from other types of mobility precisely, involuntary flows evade attempts at defining them in precise terms. Of course, many definitions related to specific populations of interest exist, beginning with the UN designation of a refugee, following the 1951 Convention and the 1967 Protocol, as someone who:

> "owing to well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion, is outside the country of his [*sic!*] nationality and is unable or, owing to such fear, is unwilling to avail himself of the protection of that country; or who, not having a nationality and being outside the country of his former habitual residence as a result of such events, is unable or, owing to such fear, is unwilling to return to it." (UNHCR, 1951/1967; Art. 1 A (2))

The UN definition is relatively narrow, being restricted to people formally recognised as refugees under international humanitarian law, even though the explicit inclusion of the notion of *fear* can help better conceptualise violence-induced migration (Kok, 2016). Broader definitions, such as those of forced displacement, range from more to less restrictive; for example, according to the World Bank:

> "forcibly displaced people [include] refugees, internally displaced persons and asylum seekers who have fled their homes to escape violence, conflict and persecution" (World Bank; http://www.worldbank.org/en/topic/forced-displacement, as of 1 September 2021).

On the other hand, the following definition of the International Association for the Study of Forced Migration (IASFM), characterises forced migrations very broadly, as:

> "Movements of refugees and internally displaced people (displaced by conflicts) as well as people displaced by natural or environmental disasters, chemical or nuclear disasters, famine, or development projects" (after Forced Migration Review; https://www.fmreview.org, as of 1 September 2021).

In several instances, pragmatic solutions are needed, so that the definition is actually determined by what can be measured, or what can be subsequently used for operational purposes by the users of the ensuing analysis. The same principle can hold for the drivers of migration and how they can be operationalised. In that spirit, Bijak et al. (2017) defined asylum-related migration as follows:

> "Asylum-related migration has therefore to jointly meet two criteria: first, it needs to be international in nature, and second, it has to be – or claimed to be – related to forced displacement, defined as forced migration due to persecution, armed conflict, violence, or violations of human rights" (Bijak et al., 2017, p.8).

This definition excludes internally displaced persons, and migrants forced to move for environment- or development-related reasons. It was also purely driven by the operational needs of the European asylum system, which was the intended user of the related analysis. For similar reasons, we use the term 'asylum migration' throughout this book, as most closely aligned with the substantive research questions that we aim to study through the lens of the model-based approach. To that end, the focus of our modelling efforts, and their possible practical applications, is on understanding the dynamics of the actual flows of people, irrespective of their legal status or specific individual circumstances.

More generally, even if a common definition could be adopted, at the higher, conceptual level, the dichotomy between forced and voluntary migration seems to some extent obsolete and not entirely valid. This is mainly attributed to the presence of a multitude of migration motives operating at the same time for a single migrant (King, 2002; Foresight, 2011; Erdal & Oeppen, 2018). The uncertainty of asylum migration is additionally exacerbated by a lack of common theoretical and explanatory framework. The aforementioned theoretical paucity of migration studies in general does not help (Arango, 2000), and the situation with respect to asylum migration is similarly problematic. Besides, in the contemporary literature there is vast disconnect between migration and refugee studies, which utilise different

theoretical approaches and do not share many common insights (FitzGerald, 2015). Comprehensive theoretical treatment of different types of migration on the voluntary-forced spectrum is rare; with examples including the important work by Zolberg (1989).

One pragmatic solution can be to focus on various factors and drivers of migration, an approach systematised in the classical *push-pull* framework of Everett Lee (1966), and since extended by many authors, including Arango (2000), Carling and Collins (2018), or Van Hear et al. (2018). Specifically in the context of forced migration, Öberg (1996) mentioned the importance of 'hard factors', such as conflict, famine, persecution or disasters, pushing involuntary migrants out from their places of residence, and leading to resulting migration flows being less self-selected. A contemporary review of factors and drivers of asylum-related migration was published in the EASO (2016) report, while a range of economic aspects of asylum were reviewed by Suriyakumaran and Tamura (2016).

In addition, uncertainty of asylum migration measurement includes many idiosyncratic features, besides those common with other forms of mobility. In particular, focus on counting administrative events rather than people results in limited information being available on the context and on migration processes themselves (Singleton, 2016). As a result, on the one hand, some estimates include duplications of the records related to the same persons; while on the other hand, some of the flows are at the same time undercounted due to their clandestine nature (*idem*).

The politicisation of asylum statistics, and their uses and misuses to fit with any particular political agenda, are other important reasons for being cautious when interpreting the numbers of asylum migrants (Bakewell, 1999; Crisp, 1999). Contemporary attempts to overcome some of the measurement issues are currently undertaken through increasing use of biometric techniques, such as the EURODAC system in the European Union (Singleton, 2016), as well as through experimental work with new data, such as mobile phone records or 'digital footprints' of social media usage (Hughes et al., 2016). This results in a patchwork of sources covering different aspects of the flows under study, as illustrated in Chap. 4 on the example of Syrian migration to Europe.

Despite these very high levels of uncertainty, formal quantitative modelling of various forms of asylum-related migration remains very much needed. Its key uses are both longer-term policy design, as well as short-term operational planning, including direct humanitarian responses to crises, provision of food, water, shelter and basic aid. In this context, decisions under such high levels of uncertainty require the presence of contingency plans and flexibility, in order to improve resilience of the migration policies and operational management systems. This perspective, in turn, requires new analytical approaches, the development of which coincides with a period of self-reflection on the theoretical state of demography, or broader population studies, in the face of uncertainty (Burch, 2018). These developments are therefore very much in line with the direction of changes of the main aims of demographic enquiries over the past decades, which are briefly summarised next.

## 2.3   Shifting Paradigm: Description, Prediction, Explanation

To trace the changes in demographic thinking about the notion of uncertainty, we need to go back to the very inception of the discipline in the seventeenth century, notionally marked by the publication of John Graunt's *Bills of Mortality* in 1662. From the outset, demography had an uneasy relationship with uncertainty and, by extension, with probability theory and statistics (Courgeau, 2012). Following a few early examples of probabilistic studies of the features of populations, the nineteenth century and the increased reliance on population censuses brought about the dominance of descriptive, and largely deterministic approaches. In that period, the questions of variation and uncertainty were largely swept under the carpet (*idem*).

Similarly, the proliferation of survey methods and data in the second half of the twentieth century offered some simple explanations of demographic phenomena in terms of statistical relationships, which still remained largely descriptive, and were missing the mechanisms actually driving population change (Courgeau et al., 2016; Burch, 2018). Only recently, especially since the 1970s and 1980s, has statistical demography begun to flourish, including a range of methods and models that apply the Bayesian paradigm, and put uncertainty at the centre of population enquiries, in such areas as prediction, small area estimation, or complex and highly-structured problems (Bijak & Bryant, 2016).

Population predictions, with their inherent uncertainty, are contemporarily seen as one of the bestselling products of population sciences (Xie, 2000). In assessing their analytical potential, Keyfitz (1972, 1981) put a reasonable horizon of population predictions at one generation ahead at most, which is already quite long, especially in comparison with other socio-economic phenomena. Within that period, the newly-born generations have not yet entered the main reproductive ages. The cohort-component mechanism of population renewal additionally ensures the relatively high levels of predictability at the population level (Lutz, 2012; Willekens, 2018): most people who will be present in a given population one generation ahead are already there.

What can reduce the predictability of population, especially in the short term, is migration, the predictive horizon of which is much shorter (Bijak & Wiśniowski, 2010), unless it is described and modelled at a very high level of generality, with very low-frequency data (Azose & Raftery, 2015). The migration uncertainty is also age-selective, affecting the more mobile age groups, such as people in the early stages of their labour market activity, more than others. This uncertainty is further amplified from generation to generation, through secondary impacts of migration on fertility and mortality rates, and through changes in the composition of populations in both origin and destination countries (for an example related to Europe, see Bijak et al., 2007).

The unpredictability of migration compounds two types of uncertainty: *epistemic*, related to imperfect knowledge, and *aleatory*, inherent to any future events, especially for complex social systems (for a detailed discussion, see Bijak & Czaika, 2020). Some migration flows are more uncertain than others, and require different

analytical tools and different assumptions on their statistical properties, such as stationarity. For some processes, or over longer horizons, coherent scenarios seem to be the only reliable way of scanning the possible future pathways (see Nico Keilman's contribution to Willekens, 1990: 42–44; echoed by Bijak, 2010). Ideally, such scenarios should be equipped with solid micro-level foundations and connect different levels of analysis, from micro (individuals), to macro (populations).

Another way to describe the uncertainty of migration flows is offered by the risk management framework, with uncertainty or volatility of a specific migration type juxtaposed against its possible societal impact (Bijak et al., 2019). Under this framework, return migration of nationals is typically less volatile – and has smaller political or societal impact – than for example labour immigration of non-nationals. Seen through the lens of risk management, the violence-induced migration, including large flows of asylum seekers, refugees and displaced persons, is typically one of the most uncertain forms of mobility, also characterised by the highest societal impact (for a conceptual overview aimed at improving forecasts, see also Kok, 2016). For such highly unpredictable types of migration, early warning models may offer some predictive insights over very short horizons (Napierała et al., 2021).

Besides, despite the advances in statistical modelling, formal description and interpretation of uncertain demographic phenomena, one key epistemological gap in contemporary demography remains: the lack of explanation of the related processes, which can be especially well seen in the studies of migration. Particularly missing are solid theoretical foundations underlying the macro-level processes (see for example Burch, 2003, 2018). Numerous micro-level studies based on surveys exist, but they do not deal with the *behaviour* of individuals, only with its observable and measurable outcomes. Even the prevailing event-history and multi-level statistical studies do not offer *causal explanations* of the mechanisms driving demographic change (Courgeau et al., 2016).

In mainstream population sciences, the discussion of micro-foundations of macro-level processes has been so far very limited. Even though the importance of explicit modelling of micro-level behaviour of individuals has been acknowledged in a few pioneering studies, such as the landmark volume by Billari and Prskawetz (2003) and its intellectual descendants and follow-ups (Billari et al., 2006; van Bavel & Grow, 2016; Silverman, 2018), the associated demographic agent-based models are still in their infancy, and their theory-building and thus explanatory potential has not yet been fully accomplished, as documented in Chap. 3 on the example of migration modelling.

At the same time, various types of computational simulation models have been gaining prominence in population studies since the beginning of the twenty-first century (Axtell et al., 2002; Billari & Prskawetz, 2003; Zaidi et al., 2009; Bélanger & Sabourin, 2017), and research on the applications of computational modelling approaches to population problems is currently gaining momentum (van Bavel & Grow, 2016; Silverman, 2018). This is because computer-based simulations, such as agent-based or microsimulation models, offer population scientists many new and exciting research possibilities. At the same time, demography remains a strongly

empirical area of social sciences, with many policy implications (Morgan & Lynch, 2001), for which computational models can offer attractive analytical tools.

So far, the empirical slant has constituted one of the key strengths of demography as a discipline of social sciences; however, there is increasing concern about the lack of theories explaining the population phenomena of interest (Burch, 2003, 2018). This problem is particularly acute in the case of the micro-foundations of demography being largely disconnected from the macro-level population processes (Billari, 2015). The quest for micro-foundations, ensuring links across different levels of the problem, thus becomes one of the key theoretical and methodological challenges of contemporary demography and population sciences.

## 2.4   Towards Micro-foundations in Migration Modelling

In order to be realistic and robust, migration (or, more broadly, population) theories and scenarios need to be grounded in solid micro-foundations. Still, in the uncertain and messy social reality, especially for processes as complex as migration, the modelling of micro-foundations of human behaviour has its natural limits. In economics, Frydman and Goldberg (2007) argued that such micro-foundations may merely involve a qualitative description of tendencies, rather than any quantitative predictions. Besides, even in the best-designed theoretical framework, there is always some residual, irreducible aleatory uncertainty. Assessing and managing this uncertainty is crucial in all social areas, but especially so in the studies of migration, given its volatility, impact and political salience (Disney et al., 2015).

In other disciplines, such as in economics, the acknowledgement of the role of micro-foundations has been present at least since the *Lucas critique* of macroeconomic models, whereby conscious actions of economic agents invalidate predictions made at the macro (population) level (Lucas, 1976). The related methodological debate has flourished for over at least four decades (Weintraub, 1977; Frydman & Goldberg, 2007). The response of economic modelling to the Lucas critique largely involved building large theoretical models, such as those belonging to the Dynamic Stochastic General Equilibrium (DSGE) class, which would span different levels of analysis, micro – individuals – as well as macro – populations (see e.g. Frydman & Goldberg, 2007 for a broad theoretical discussion, and Barker & Bijak, 2020 for a specific migration-related overview).

Existing migration studies offer just a few overarching approaches with a potential to combine the micro and macro-level perspectives: from multi-level models, that belong to the state of the art in statistical demography (Courgeau, 2007), to conceptual frameworks that potentially encompass micro-level as well as macro-level migration factors. The key examples of the latter include the push and pull migration factors (Lee, 1966), with recent modifications, such as the push-pull-plus framework (Van Hear et al., 2018), and the value-expectancy model of De Jong and Fawcett (1981). In the approach that we propose in this book, however, the link between the different levels of analysis is of statistical and computational nature,

rather than being analytical or conceptual. In particular, in our approach, bridging the gap between the different levels of analysis involves building micro-level simulation models of migration behaviour, which can then be calibrated to some aspects of macro-level data.

One alternative approach for combining different levels of analysis involves building microsimulation models, whereby simulated individuals are subject to transitions between different *states* according to empirically derived rates, which are typically data-driven (Zaidi et al., 2009; Bélanger & Sabourin, 2017). Such models can be limited by the availability of detailed data, and often follow simple assumptions on the underlying mechanisms, for example Markovian 'lack of memory' (Courgeau et al., 2016). In contrast, agent-based models, based on interacting individual agents, allow for explicit inclusion of feedback effects and modelling the bidirectional impact of macro-level environment on individual behaviour and vice versa through the 'reverse causality' mechanisms (Lorenz, 2009). Still, it is recognised that many of the existing agent-based attempts are too often based on unverifiable assumptions and axioms (Conte et al., 2012).

Agent-based models focus on representing the behaviour of simulated individuals – agents – in artificial computer simulations, through applying micro-level behavioural rules to study the resulting patters emerging at the macro level. Such models, while not predictive *per se*, can be used for a variety of objectives. Epstein (2008) identified sixteen aims of modelling, from explanation, to guiding data collection, studying the range of possible outcomes, and engagement with the public. The perspective of generating explanatory mechanisms for migration through simulations and model-building, and enabling experimentation in controlled conditions *in silico*, are both very appealing to demographers (Billari & Prskawetz, 2003), and potentially also to the users of their models, including policy makers. We explore many of these aspects throughout this book.

Given the state of the art of demographic modelling, important methodological advances can be therefore achieved by building agent-based simulation models of international migration, combined in a common framework with the recent cutting-edge developments across a range of disciplines, including demography, statistics and experimental design, computer science, and cognitive psychology, the latter shedding light on the specific aspects of human decision making. This approach can enhance the traditional demographic modelling of population-level dynamics by including realistic and cognitively plausible micro-foundations.

There are several important examples of work which look at applications of agent-based modelling to social science, beginning with the seminal work of Schelling (1971, 1978). More recently, a specialised field of social simulation has emerged (Epstein & Axtell, 1996; Gilbert & Tierna, 2000), as has the analytical sociology research programme (Hedström & Swedberg, 1998; Hedström, 2005). Recently, the topic was explored, and the field thoroughly reviewed by Silverman (2018). As mentioned above, the pioneering demographic book advocating the use of agent-based models (Billari & Prskawetz, 2003) was followed by subsequent extensions and updates (e.g. Billari et al., 2006; van Bavel & Grow, 2016). In parallel, microsimulation models have been developed and extensively applied (for an

overview, see e.g. Zaidi et al., 2009; Bélanger & Sabourin, 2017). In migration research, several examples of constructing agent-based models exist, such as Kniveton et al. (2011) or Klabunde et al. (2017), with a more detailed survey of such models offered in Chap. 3.

In general, agent-based models have complex and non-linear structures, which prohibit a direct analysis of their outcome uncertainty. Promising methods which could enable indirect analysis include Gaussian process (GP) emulators or meta-models – statistical models of the underlying computational models (Kennedy & O'Hagan, 2001; Oakley & O'Hagan, 2002), or the Bayesian melding approach (Poole & Raftery, 2000), implemented in agent-based transportation simulations (Ševčíková et al., 2007). In demography, prototype GP emulators have been tested on agent-based models of marriage and fertility (Bijak et al., 2013; Hilton & Bijak, 2016). A general framework for their implementation is that of (Bayesian) statistical experimental design (Chaloner & Verdinelli, 1995), with other approaches that can be used for estimating agent-based models including, for example, Approximate Bayesian Computations (Grazzini et al., 2017). A detailed discussion, review and assessment of such methods follows in Chap. 5.

Before embarking on the modelling work, it is worth ensuring that the outcomes – models – have realistic potential for increasing our knowledge and understanding of demographic processes. The discussion about relationship between modelling and the main tenets of the scientific method remains open. To that end, we discuss the epistemological foundations of model-based approaches next, with focus on the question of the origins of knowledge in formal modelling.

## 2.5  Philosophical Foundations: Inductive, Deductive and Abductive Approaches

There are several different ways of carrying out scientific inference and generating new knowledge. The deductive reasoning has been developed through millennia, from classical syllogisms, whereby the conclusions are already logically entailed in the premises, to the hypothetico-deductive scientific method of Karl Popper (1935/1959), whereby hypotheses can be falsified by non-conforming data. The deductive approaches strongly rely on hypotheses, which are dismissed by the proponents of the inductive approaches due to their arbitrary nature (Courgeau et al., 2016).

The classical inductive reasoning, in turn, which underpins the philosophical foundations of the modern scientific method, dates back to Francis Bacon (1620). It relies on inducing the formal *principles* governing the processes or phenomena of interest (Courgeau et al., 2016), at several different levels of explanation. These principles, in turn, help identify the key *functions* of the processes or phenomena, which are required for these processes or phenomena to occur, and to take such form as they have. The identified functions then guide the observation of the empirical

properties, so that in effect, the observed variables describing these properties can illuminate the *functional structures* of the processes or phenomena as well as the *functional mechanisms* that underpin them.[1]

When it comes to hypotheses, the main problem seems to be not so much their existence, but their haphazard and often not properly justified provenance. To help address this criticism, a third, less-known way of making scientific inference has been proposed: *abduction*, also referred to as 'inference to the best explanation'. The idea dates back to the work of Charles S. Peirce (1878/2014), an American philosopher of science working in the second half of the nineteenth century and the early twentieth century. His new, pragmatic way of making a philosophical argument can be defined as "inference from the body of data to an explaining hypothesis" (Burks, 1946: 301).

Seen in that way, abduction appears as a first phase in the process of scientific discovery, with setting up a novel hypothesis (Burks, 1946), whereas deduction allows subsequently for deriving testable consequences, while modern induction allows their testing, for example through statistical inference. As an alternative classification, Lipton (1991) labelled abduction as a separate form of inductive reasoning, offering 'vertical inference' (*idem*: 69) from observable data to unobservable explanations (theory), allowing for the process of discovery. The consequences of the latter can subsequently follow deductively (*idem*). Thanks to the construction and properties of abductive reasoning, this perspective has found significant following within the social simulation literature, to the point of equating the methods with the underpinning epistemology. To that end, Lorenz (2009: 144) explicitly stated that "simulation model is an abductive process".

Some interpretations of abductive reasoning stress the pivotal role it plays in the sequential nature of the scientific method, as the stage where new scientific ideas come from in a process of creativity. At the core of the abductive process is surprise: observing a surprising result leads to inferring the hypothesis that could have led to its emergence. In this way, the (prior) beliefs, confronted by a surprise, lead to doubt and enable further, creative inference (Burks, 1946; Nubiola, 2005), which in itself has some conceptual parallels with the mechanism of Bayesian statistical knowledge updating.

There is a philosophical debate as to whether the emergence of model properties as such is of ontological or epistemological nature. In other words, whether modelling can generate new facts, or rather help uncover the patterns through improved knowledge about the mechanisms and processes (Frank et al., 2009). The latter interpretation is less restrictive and more pragmatic (*idem*), and thus seems better suited for social applications. As an example, in demography, a link between discovery (surprise) and inference (explanation) was recently established and

---

[1] The notion of classical induction is different from the concept of induction as developed for example by John Stuart Mill, where observables are generalised into conclusions, by eliminating those that do not aid the understanding of the processes under study, for example in the process of experimenting (Jacobs 1991). The two types of induction should not be confused. On this point, I am indebted to Robert Franck and Daniel Courgeau for detailed philosophical explanations.

formalised by Billari (2015), who argued that the act of discovery typically occurs at the population (macro) level, but explanation additionally needs to include individual (micro)-level foundations.

Abduction, as 'inference to the best explanation', is also a very pragmatic way of carrying out the inferential reasoning (Lipton, 1991/2004). What is meant by the 'best explanation' can have different interpretations, though. First, it can be the best of the candidate explanations of the probable or approximate truth. Second, it can be subject to an additional condition that the selected hypothesis is satisfactory or 'good enough'. Third, it can be such an explanation, which is 'closer to the truth' than the alternatives (Douven, 2017).

The limitations of all these definitions are chiefly linked to a precise definition of the criterion for optimality in the first case, satisfactory quality criteria in the second, as well as relative quality and the space of candidate explanations in the third. One important consideration here is the parsimony of explanation – the Ockham's razor principle would suggest preferring simple explanations to more complex ones, as long as they remain satisfactory. Another open question is which of these three alternative definitions, if any, are actually used in human reasoning (Douven, 2017)?

In any case, a lack of a single and unambiguous answer points out to lack of strict identifiability of abductive solutions to particular inferential problems: under different considerations, many candidate explanations can be admissible, or even optimal. This ambiguity is the price that needs to be paid for creativity and discovery. As pointed out by Lorenz (2009), abductive reasoning bears the risk of an *abductive fallacy*: given that abductive explanations are sufficient, but not necessary, the choice of a particular methodology or a specific model can be incorrect.

These considerations have been elaborated in detail in the philosophy of science literature. In his comprehensive treatment of the approach, Lipton (1991/2004) reiterated the pragmatic nature of inference to the best explanation, and made a distinction between two types of reasoning: 'likeliest', being the most probable, and 'loveliest', offering the most understanding. The former interpretation has clear links with the probabilistic reasoning (Nubiola, 2005), and in particular, with Bayes's theorem (Lipton, 2004; Douven, 2017). This is why abduction and Bayesian inference can be even seen to be 'broadly compatible' (Lipton, 2004: 120), as long as the elements of the statistical model (priors and likelihoods) are chosen based on how well they can be thought to explain the phenomena and processes under study. In relation to the discussion of psychological realism of the models of human reasoning and decision making (e.g. Tversky & Kahneman, 1974, 1992), formal Bayesian reasoning can offer rationality constraints for the heuristics used for updating beliefs (Lipton, 2004).

There are important implications of these philosophical discussions both for modelling, as well as for practical and policy applications. To that end, Brenner and Werker (2009) argued that simulation models built by following the abductive principles at least partially have a potential to reduce the error and uncertainty in the outcome. In particular, looking at the modelled structures of the policy or practical problem can help safeguard against at least some of the unintended and undesirable consequences (*idem*), especially when they can be identified through departures from rationality.

In that respect, to help models achieve their full potential, the different philosophical perspectives need to be ideally combined. As deduction on its own relies on assumptions, induction implies uncertainty, and abduction does not produce uniquely identifiable results, the three perspectives should be employed jointly, although even then, uncertainty cannot be expected to disappear (Lipton, 2004; Brenner & Werker, 2009). These considerations are reflected in the nascent research programme for model-based demography, the main tenets of which we discuss in turn.

## 2.6 Model-Based Demography as a Research Programme

The methodology we propose throughout the book is inspired by the principles of the *model-based research programme* for demography, recently outlined by Courgeau et al. (2016), who were inspired by Franck (2002). In parallel, similar propositions have been developed by other prominent authors, such as Burch (2018), in a tradition dating back to Keyfitz (1971). Among the different approaches to demographic modelling, Courgeau et al. (2016) suggested that the model-building process should follow the classical inductive principles from the bottom up. In this way, the process should start by observing the key population properties generated by the process under study (migration), followed by inferring the *functional structures* of these processes in their particular context, identifying the relevant variables, and finally conceptual and computational modelling. The results of the modelling should allow for identifying gaps in current knowledge and provide guidance on further data collection. By so doing, the process can be iterated as needed, as argued by Courgeau et al. (2016), ideally following the broad principles of classical inductive reasoning.

It is worth stressing that the proposed model-based programme is not the same as an approach that relies purely on agent-based modelling. First, the model-based approaches can involve different types of models: agent-based ones are an obvious possibility, but microsimulations or formal mathematical models can also be used, alongside the statistical models used to unravel the properties of analytical or computational models they are meant to analyse. Second, as argued in Chap. 3, agent-based models alone, especially those applied to social processes such as migration, necessarily have to make many arbitrary and *ad hoc* assumptions, unless they can be augmented with additional information from other sources – observations, experiments, and so on – as proposed in the full model-based approach advocated here. From that point of view, the model-based approach includes a (computational or analytical) model at its core, but goes beyond that – and the *process* of arriving at the final form of the model is also much more involved than the programming of a model alone.

The existing agent-based attempts at describing migration, reviewed and evaluated in more detail in Chap. 3, offer a good starting point for the model-building process. In particular, Klabunde et al. (2015) looked at the staged nature of the

decision process, following the Theory of Planned Behaviour (Ajzen, 1985), whereby behaviour results from intentions, formed on the basis of beliefs, norms and attitudes, and moderated by actual behavioural control. None of the existing approaches, however, explicitly represent key cognitive aspects of decision-making mechanisms, nor do they include a comprehensive uncertainty assessment at the different levels of analysis. Our proposed model-based approach offers insights into bottom-up modelling based on a range of information sources, addressing some of the key epistemological limitations of simulations, especially of human decisions.

There are many other building blocks that can facilitate modelling: importantly, despite high uncertainty, migration is characterised by stable regularities in terms of its spatial structures (Rogers et al., 2010) and age profiles (Rogers & Castro, 1981). The latter is an outcome of links with life course and other demographic processes, such as family formation or childbearing (Courgeau, 1985; Kulu & Milevski, 2007). The role of migrant networks in the perpetuation of migration processes is also well recognised (Kritz et al., 1992; Lazega & Snijders, 2016). For such elements – networks and *linked lives* – agent-based models are a natural tool of scientific enquiry (Noble et al., 2012). Following the general philosophy of Ben-Akiva et al. (2012), it is also worthwhile distinguishing the *process* of migration decision making at the individual level, and the *context* at the group and societal levels, integrated within a common multi-level analytical model. A joint modelling of different levels of analysis was also suggested in the *Manifesto of computational social science* by Conte et al. (2012). In the same work, Conte et al. (2012) suggested that computational social science modelling should be more open to non-traditional sources of data, beyond surveys and registers, and in particular embrace tailor-made experimentation under controlled conditions.

Many of these different elements are used in the application of the model-based approach presented throughout this book. The empirical experiments focus on different aspects of human decision-making processes, such as choices between different options (Ben-Akiva et al., 2012), the role of uncertainty – especially the subjective probabilities and possible biases – as well as attitudes to risk (Gray et al., 2017), which are discussed in more detail in Chap. 6. In this way, the purpose of a scientific enquiry becomes as much about the model and the related analysis, as it is about the process of the iterative improvement of the analytical tools and an increase in their sophistication. In philosophical terms, the proposed approach also addresses the methodological suggestions made by Conte et al. (2012) that different types of empirical data are used throughout the model construction process, not merely for final validation, which is understood here as ensuring alignment between the model and some aspects of the observed reality.

Nevertheless, one important challenge of designing and implementing such a modelling process remains: how to combine simulations with other analytical methods, including statistics, as well as experiments, with a strong empirical base (Frank et al., 2009)? To that end, Courgeau et al. (2016) stressed the role of appropriate experimental design and related statistical methods to bring the different methodological threads together, and to align model-based enquiries closer with the classical inductive scientific research programme, dating back to Francis Bacon (1620;

after: *idem*). The broad tenets of this approach are followed throughout this book, and its individual components are presented in Part II.

In the model-based programme, as proposed by Courgeau et al. (2016), the objective of modelling is to *infer* the functional structures that generate the observed social properties. Here, the empirical observables are necessary, but not sufficient elements in the process of scientific discovery, given that for any set of observables, there can be a range of non-implausible models generating matching outcomes (*idem*). At the same time, as noted by Brenner and Werker (2009), the modelling process needs to explicitly recognise that the errors in inference are inevitable, but modellers should aim to reduce them as much as possible.

In what can be seen as a practical solution for implementing a version of the model-based programme, Brenner and Werker (2009:3.6) advocated four steps of the modelling process:

(1)  Setting up the model based on all available empirical knowledge, starting from a simple variant, and allowing for free parameters, wherever data are not available (abduction);
(2)  Running the model and calibrating it against the empirical data for some chosen outputs, excluding the implausible ranges of the parameter space (induction, in the modern sense);
(3)  On that basis, classifying observations into classes, enabling alignment of theoretical explanations implied by the model structure with empirical observations (another abduction);
(4)  Use of the calibrated model for scenario and policy analysis (which *per se* is a deductive exercise, notwithstanding the abductive interpretation given by Brenner & Werker, 2009).

In this way, the key elements of the model-based programme become explicitly embedded in a wider framework for model-based policy advice, which makes full use of three different types of reasoning – inductive, abductive and deductive – at three different stages of the process. Additionally, the process can implicitly involve two important checks – *verification* of consistency of the computer code with the conceptual model, and *validation* of the modelling results against the observed social phenomena (see David, 2009 for a broad discussion).

As a compromise between the ideal, fully inductive model-based programme advocated by Courgeau et al. (2016) and the above guidance by Brenner and Werker (2009), we propose a pragmatic variant of the model-based approach, which is summarised in Fig. 2.1. The modelling process starts by defining the specific research question or policy challenge that needs explaining – the model needs to be specific to the research aims and domain (Gilbert & Ahrweiler, 2009, see also Chap. 3). These choices subsequently guide the collection of information on the properties of the constituent parts of the problem. The model construction then ideally follows the classical inductive principles, where the functional structure of the problem, the contributing factors, mechanisms and the conceptual model are inferred. If a fully inductive approach is not feasible, the abductive reasoning to provide the 'best explanation' of the processes of interest can offer a pragmatic alternative.

Subsequently, the model, once built, is internally verified, implemented and executed, and the results are then validated by aligning them with observations. This step can be seen as a continuation of the inductive process of discovery. The nature of the contributing functions, structures and mechanisms is unravelled, by identifying those elements of the modelled processes without which those processes would
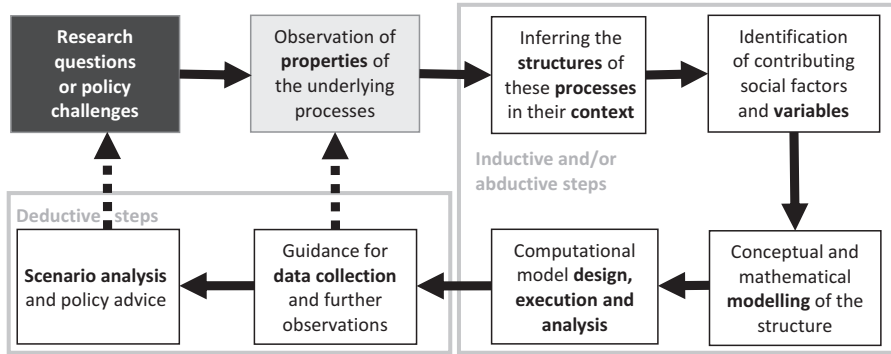
**Fig. 2.1** Basic elements of the model-based research programme. (Source: own elaboration based on Courgeau et al., 2016: 43, and Brenner and Werker, 2009)

not occur, or would manifest themselves in a different form. At this stage, the model can also help identify (deduce) the areas for further data collection, which would lead to subsequent model refinements. At the same time, also in a deductive manner, the model generates derived scenarios, which can serve as input to policy advice. These scenarios can give grounds to new or amended research or policy questions, at which point the process can be repeated (Fig. 2.1).

Models obtained by applying the above principles can therefore both enable scenario analysis and help predict structural features and outcomes of various policy scenarios. The model outcomes, in an obvious way, depend on empirical inputs, with Brenner and Werker (2009) having highlighted some important pragmatic trade-offs, for example between validity of results and availability of resources, including research time and empirical data. These pragmatic concerns point to the need for initiating the modelling process by defining the research problem, then building a simple model, as a first-order approximation of the reality to guide intuition and further data collection, followed by creating a full descriptive and empirically grounded version of the model.

At a more general level, modelling can be located on a continuum from general (nomological) approaches (Hempel, 1962), aimed at uncovering idealised laws, theories and regularities, to specific, unique and descriptive (ideographic) ones (Gilbert & Ahrweiler, 2009). The blueprint for modelling proposed in this book aims to help scan at least a segment of this conceptual spectrum for analysing the research problem at hand.

In epistemological terms, the guiding principles of the abductive reasoning can be seen as a pragmatic approximation of a fully inductive process of scientific enquiry, which is difficult whenever our knowledge about the functions, structures and mechanisms is limited, incomplete, poor quality, or even completely missing. In the context of social phenomena, such as migration, these limitations are paramount. This is why the approach adopted throughout the book sees the classical induction as the ideal philosophy to underpin model-based enquiries, and the abductive reasoning as a possible real-life placeholder for some specific aspects. In this way, we

aim to offer a pragmatic way of instantiating the model-based research programme in such situations, where applying the fully inductive approach for every element of the modelling endeavour is not feasible. We discuss the elements of the proposed methodology in more detail in Part II.