Jakub Bijak, Mathias Czaika

Assessing Uncertain Migration Futures: A Typology of the Unknown

Deliverable 1.1

QuantMig has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870299.

History of changes

Version	Date	Changes
1.0	5 June 2020	Issued for Consortium Review
1.1	30 June 2020	First version submitted as official deliverable to the EC

Suggested citation

Bijak J and Czaika M (2020) Assessing Uncertain Migration Futures – A Typology of the Unknown. QuantMig Project Deliverable D1.1. Southampton / Krems: University of Southampton and Danube University Krems.

Dissemination level

PU Public

Key words

Complexity, Migration drivers, Migration processes, Prediction, Uncertainty

Acknowledgments

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870299 QuantMig: Quantifying Migration Scenarios for Better Policy. We are very grateful to Marta Bivand Erdal, Ketevan Glonti, Nico Keilman, Giampaolo Lanzieri, Constantinos Melachrinos, Akira Soto-Nishimura, Simone Russo and Frans Willekens, for their comments that helped improve an earlier draft. Needless to say, all the remaining errors and inaccuracies remain ours. This document reflects the authors' view and the Research Executive Agency of the European Commission are not responsible for any use that may be made of the information it contains.

Cover photo: iStockphoto.com/Guenter Guni

Table of Contents

List of Tables

Assessing Uncertain Migration Futures: A Typology of the Unknown

QuantMig Deliverable D1.1

Jakub Bijak[a]✉ and Mathias Czaika[b]✉

[a] University of Southampton, UK; [b] Danube University Krems, Austria;
✉ Corresponding authors: j.bijak@soton.ac.uk; mathias.czaika@donau-uni.ac.at

Abstract

Migration is one of the most, if not the most uncertain of the major demographic processes. A complex network of interacting drivers and factors, coupled with a key role of human agency in migration decisions, well-known issues with ambiguous conceptualization, and problematic measurement of migration flows, all contribute to the very high levels of complexity and uncertainty. This limits our ability to study migration, not to mention predicting it. The aim of this paper is to shed light on the different facets of migration uncertainty, by offering a unique typology of the knowable and unknowable features of migration, with a special focus on the complexity of the underlying drivers. The typology is centered around the limits of knowledge of different migration concepts, definitions, drivers and multidimensional driver complexes, measurements, features, spatio-temporal and life course regularities, as well as possible impacts. By doing so, we hope to illuminate the promising areas of migration studies which have the highest potential of reducing the associated uncertainty through further research in the future. At the same time, the typology helps to identify those features of migration processes, for which the uncertainty is irreducible, and needs to be adequately managed. The conceptual work underpinning the typology is based on a systematic review of the state of the art in migration studies with respect to how migration uncertainty and complexity are being defined, measured, and analysed.

1. Introduction

Policymakers, governmental and non-governmental institutions, economic actors, and the general public exhibit an ever-increasing appetite for anticipating and preparing for future migration flows, both on their own, as direct input into policy decisions, as well as one of the components of population projections and forecasts. In 2020, five years after the height of the often-called European migration crisis and its associated societal, political and discursive disruptions, various stakeholders ask scientists to provide better predictions of future migration.

It is worth noting the semantic difference between different terms related to assessing various possible migration futures[1]. Despite slight semantic differences between *forecasts* and *predictions*, suggested by some authors, in this paper we will use these terms interchangeably as unconditional statements about the future. This contrasts with *projections* or *scenarios*, which are conditional upon their assumptions and underlying narratives (Bijak 2010). A migration *forecast* aims therefore to provide the most plausible development of the future migration size and composition, which can be probabilistic, while *scenarios* can contain 'what-if' alternatives, based on the narrative stories and other related assumptions (Bijak 2010; de Beer 2011). In the demographic literature, scenarios are often associated with projection variants, and can be used for planning for desired developments under certain assumptions (de Beer 2011).

Similar semantic ambiguities relate to the definition of *migration* and a *migrant*. The typical definitions, which are mirrored in the official population statistics, involve events related to people changing their place or residence (usual or legal) for a certain time period (for an overview, see Lanzieri 2019a, 2019b). In this paper, we mostly focus on *international migration*, related with crossing national boundaries, although many of the points we make are general and apply equally to *internal migration* and other forms of mobility, not to mention that there can be some equivocation in classifying migration into either form[2]. As we argue in the paper, many other aspects the definitions applied are far from being unambiguous as well.

Even though working for more than a hundred years on a theoretical and empirical comprehension of migration processes (see e.g. Ravenstein 1885, and Rees and Lomax 2020), migration scholars are still not able to predict *accurate* numbers of migrants, regardless of circumstances – which is true for most, if not all forecasts, especially in the social domain. The fundamental epistemological question is, to what extent migration processes are predictable at all, even exact to providing a probability distribution of possible future outcomes. This question links with another one, which is a thread underlying this paper: which types of uncertainty in studying migration are *epistemic* – related to imperfect knowledge, which can potentially be obtained through more research, rendering these aspects of migration processes at least potentially *knowable*, and which ones are *aleatory* – fundamentally irreducible, related to unpredictable shocks and their impacts, and thus in principle *unknowable*. In other words, the epistemic uncertainty represents our ignorance about the processes, and aleatory reflects their intrinsic randomness. We are fully aware of the difficulty of delineating the boundary precisely[3], but are nonetheless convinced that at least making such an attempt is a necessary first step to providing a better and more realistic policy advice.

---

[1] Throughout the paper, we use the plural form ('futures') to denote the multiple possibilities across a range of time horizons in the future (singular). Likewise, the notion of 'uncertainty' refers to the key concept underlying the whole analysis, and remains singular.

[2] See, e.g. King's (2002) argument that migration within the single market of the European Union (and European Economic Area) evades a precise classification as either internal or international, sharing important characteristics with both.

[3] Consider, for instance, the migration impacts of a natural disaster, such as a tsunami. Even if the disaster itself is aleatory, the migration response is mediated by the level of government and civic preparedness, including early-warning systems, procedures, infrastructure, humanitarian assistance, and so on. The presence or absence of such factors is an epistemic feature of the ensuing migration, and is something that remains to some extent under the control of policy makers. (We thank Frans Willekens for this example.)

Historically, (net) migration forecasts have been systematically included in population projections since the UN efforts in the 1970s (Rogers 1975: 1–2), even though the cohort-component methodology for population projections existed before, and despite migration being acknowledged to be an important, yet uncertain component of population change (e.g. Holzer 1959). Since then, the prominence of standalone migration forecasting efforts has to a large extent followed the ebbs and flows of the political interest and the place of migration on the policy agenda. For example, several studies were published around the fall of the Soviet Union and the eastward enlargement of the European Union, chiefly based on econometric and statistical models, and generating substantial policy interest (for an overview, see Bijak 2010). This was later followed by academic and policy interest in the intersection between migration and environmental change (e.g. Jäger et al. 2009; Foresight 2011, De Haas et al. 2011, Ramirez et al. 2015). Currently, migration predictions and scenarios are back on the political agenda, especially after the 2015–16 crisis, with many scenario-based efforts undertaken across Europe (Frontex 2016; Acostamadiedo et al. 2020; Sardoschau 2020).

Accuracy is obviously a relative concept and while social scientists and experts have sometimes anticipated future migration trends slightly more accurately, at other times they completely failed. To that end, overall, the track record of migration projections and forecasts is not too strong. In their assessment of forecasts produced by statistical offices in 14 European countries, Keilman and Pham (2004) found that migration has been consistently underestimated in historical forecasts. Similarly, Wilson (2017) stated that large migration errors are the norm, both the long-run levels of migration and their short-term fluctuations. In addition, Shaw (2007) and Wilson (2017) argued that the official migration forecasts, respectively for the UK and Australia, for the past four decades were mostly highly inaccurate, both for individual years and cumulatively over longer periods. The reasons for that include, in varying proportions, the occurrence of shock events, unpredictable changes to migration drivers, as well as large variation in the way migration responded to these changes.

In another overview of forecasting work, Keilman (2008) concluded that recent forecasts made by European national statistical offices in the early 21st century were not more accurate in forecasting net migration than was the case in the past. One example of a country, where accurate predictions were very hard due to a dynamic population change and a small population is Ireland, which had close to zero net migration in 1995, increasing to around +105,000 in 2007 before falling again to nearly zero in 2009, at the outset of the financial crisis, turning negative in 2010, with positive values not seen until 2015 – none which was predicted by the Irish Statistical Office (Fitzgerald 2018)[4].

Historical examples of migration forecasts vary in terms of how successful they were, which also raises the question of how "success" needs to be defined in the context of a very unpredictable process. It is easier to point to examples of very unsuccessful predictions, such as those that overestimated migration from the former Soviet Union post-1991 by a large number (experts cited by Vishnevsky and Zayonchkovskaya 1994), or vastly underestimated migration into the United Kingdom after the 2004 enlargement of the European Union (Dustmann et al. 2003). Selective reviews of past forecasting efforts are offered in Bijak (2010) and Bijak et al. (2019). Given the low levels of predictability of migration, while no forecasts can be fully "successful" in other way than by fluke, some have clearly lower and better calibrated errors than others. In this way, the success

[4] With thanks to Akira Soto-Nishimura.

of forecasts can be measured in terms of their error distributions: whether the errors are well-calibrated, and how large they are (e.g. Bijak et al 2019). This perspective formally acknowledges the famous adage by George EP Box, that "all models are wrong, but some are useful".

To provide a hypothetical, perfectly accurate predication of future migration at a specific time and location, or indeed for any prediction, four key ingredients would be needed. First, there is a need for a perfect *theory and understanding* of the complex interplay of multiple migration drivers. Second, perfect *measurement and data* on multiple forms of past and current migration and its multiple drivers would need to be available over time and space. Third, we would need to have perfectly specified *models* analysing migration and driver data. Fourth, which is the most demanding ingredient, we would need to possess perfect *knowledge of the future* development of multiple drivers. This could be expected of Laplace's demon (Frigg et al. 2014), but not of human forecasters.

The degree of inaccuracy in predicting *normal forms* of migration depend therefore on knowledge of each of these four ingredients. Lack of such knowledge establishes epistemic uncertainty which we will discuss in greater detail in this paper. However, migration caused by unpredictable and singular shock events (a tsunami, a sudden conflict, or fundamental regime change) can be equally unpredictable and singular. The impossibility of predicting such singular shock events creates aleatory uncertainty regarding future migration. In addition, the sheer complexity of migration drivers and their interactions, operating across different spatial and temporal scales, compounds both types of uncertainty even further. As a result of these limitations, our view of the role of migration theory and drivers is far from being 'nomological', whereby the forecasts require precise theoretical description of how migration processes work, from which predictions could be simply deduced (Hempel 1962; Chojnicki 1970). In this paper, we settle instead for a weaker interpretation of theoretical or explanatory endeavours, as aiming to shed some light, as imprecisely as may be, on the mechanisms underpinning migration processes and their associations with other variables[5].

Uncertainty in forecasting is obviously not static but increases with the projected time horizon. While epistemic uncertainty is easier to handle for the very short term, the accuracy in predicting future migration is continuously declining the more the period of interest is long-term. Making any – even probabilistic – statements about future migration flows beyond the horizon of five to ten years becomes too risky, as the uncertainty can be overpowering (see Bijak and Wiśniowski 2010). There, a recourse to other tools, such as scenario-based projections may be needed. In this paper, we look therefore across a range of time horizons, from very short, daily or weekly, to very long, spanning years or decades, to identify the dominant uncertainty types in each case, and to propose possible methods of managing the uncertainty that would be well suited for different futures.

Even though a dynamically increasing uncertainty is relevant for all types and forms of migration, its extent can vary significantly. This leads to suggestions for making scenarios for disaggregated groups sharing some common characteristics (de Beer 2008). While family migration, for instance, is highly dependent on the volume of prior permanent immigrants from a specific context (region or community) who are subsequently joined by their family members, asylum migration or irregular forms of migration are far less path-dependent but more erratic and volatile (*idem*).

---

[5] We thank Nico Keilman for this interpretation.

A formal classification of different forms of migration in terms of their uncertainty and volatility, as well as and potential societal and policy impact, can be based on the risk management framework. In the example of UK migration, Bijak et al. (2019) proposed and a traffic-light system for assessing different migration types, from the better predictable and less consequential, such as the return flows of nationals (green), to those of highest volatility and impact, including asylum-related flows, such as those observed in Europe during the 2015–16 crisis[6]. It is worth noting here that the notion of 'impact' may have normative connotations and, depending on the perspective, focus on 'positive' or 'negative' outcomes[7]. We acknowledge that migration can bring about challenges as well as opportunities for the sending, receiving and transit countries and communities, and for individual migrants. Still, to avoid unnecessary value judgements, throughout this paper we define *migration impact* simply as 'consequence that requires action and resources on the part of decision makers.'

Finally, the level of spatial granularity also matters in determining the uncertainty of forecasts. Users of migration or population projections may require more fine-grained forecasts than those made at the national level. Subnational forecasts have less of a history and are typically not done by supranational organizations. In general, forecast error is larger at the subnational level than at the national level (Wilson and Shalley 2019). At the subnational level, forecast error generally decreases with population size but error level tends to stabilize after size threshold is reached and error tends to be lower for places with slow rates of change.

This paper aims to assess the multiple sources of uncertainty which are individually and conjointly limiting the ability to predict accurate numbers of future migrants. We focus specifically on *migration flows*, being less predictable and more volatile than *migrant stocks*, which exhibit larger inertia, driven by the mechanism of population renewal. At the same time, we look at flows as *aggregate processes*, rather than *individual moves*, even though we discuss decision making mechanisms in the review of migration drivers. Even though our analysis is motivated by international migration in Europe, the discussion is more universal: if anything, in many other parts of the world, some types of uncertainty – such as those related to data and measurement – are even more profound.

In Section 2, we propose a conceptual framework for understanding complex driver environments established by the configurational interactions of multiple migration drivers across space and time. The focus on a variety of drivers provides a starting point for the discussion of the different guises of uncertainty, to which the driver complexity contributes. In Section 3, we further elaborate on these different types of uncertainty that are limiting the ability to generate precise estimates of a specific type or form of future migration. Section 4 includes a discussion the methodological advances and limitations in incorporating epistemic and aleatory uncertainty in alternative forecasting approaches, as well as some practical and policy recommendations by way of concluding.

---

[6] At the same time, it is well recognized that some conceptual categories of migration overlap, their definitions often being driven by legal or administrative considerations – after all, asylum-related migration may be at the same time also linked to family reunion, study, or enhancing economic prospects. At the same time, other categories are strictly exclusive, for example returning nationals cannot seek asylum in their country of nationality. This complexity, coupled with the often-blurred distinction between the various dichotomies or categories of migration flows, which do not map onto clear-cut definitions, adds an additional layer of conceptual complexity (King 2002; Foresight 2011; Bivand Erdal and Oeppen 2018).

[7] We thank Marta Bivand Erdal for drawing our attention to this.

2. Conceptualizing complex migration driver environments

Why exactly do people (want to) migrate? At specific moments in people's lives, a number of factors are coming together and stimulate migration intentions and aspirations. Given some achievable opportunities, this may result in temporary or permanent migration of individuals and their families. Factors that drive migration intentions (or aspirations) and actual moves are manifold and multifaceted, and for decades migration researchers are identifying and describing factors and contexts that facilitate or constrain migration processes (Massey et al. 1993, 1999). Migration scholars studying drivers of migration are basically asking: how do migration driving factors operate in time and space? To what extent, and in which ways, do they influence, i.e. trigger or hinder, migration decision-making of some people, but not of others? And, how do multidimensional migration drivers mutually interact and create so-called *driver configurations* which may affect some people more than others in aspiring and realising migration as a viable behavioural option?

Drivers of migration have been studied for decades and the scientific literature has identified a number of fundamental dimensions of migration drivers including economic, political, social, cultural, demographic, and ecological factors (Ghatak et al. 1996; Hagen-Zanker 2008; King 2012; Massey et al. 1993, Migali et al. 2018). The circumstances, the ways and modes, and the extent to which sets of driving factors may influence individual migration decision-making and larger migration processes are dependent on the *functionality* of migration drivers, which is a central aspect in understanding the specific role of single or combinations of multiple drivers may play at different stages of a migration (decision-making) process. Migration as a behavioural option is highly context-dependent and, as such, the configuration of complex driver environments is very specific to the time and place in which migration aspirations are formed and decisions taken. However, context-specific functionalities of specific migration drivers can be generalised and categorised along some key functions (cf. Van Hear et al. 2018). *Predisposing* drivers, for instance, define fundamental societal structures and structural disparities. As the basic methodological premise, we may assume that (potential) migrants respond to extrinsic or intrinsic stimuli when deciding about migration (Czaika and Reinprecht 2020). From this perspective, predisposing factors define the broadest, most fundamental layer of *opportunity structures* (cf. de Haas 2010).

Predisposing factors, also often termed 'root causes', in particular in policy circles, are also thought of as the socio-economic, political, environmental conditions that induce departures such as poverty, repression, environmental degradation or violent conflict (Carling and Talleraas 2016). Predisposing factors do not affect directly and 'unfiltered' people's migration decision-making but are rather moderated through some *mediating* factors that facilitate, constrain, accelerate, consolidate, or diminish migration (Van Hear et al. 2018). For instance, ethno-religious fragmentations of societies in combination with socio-economic inequalities are often deeply embedded in wider societal structures. Still, they may be mediated by cultural norms (e.g. class, castes, social status) or provisions of political and civil rights, which may sometimes reduce or thwart the migration stimulating effect of structural inequalities. Thus, predisposing factors such as social, economic, political, demographic, environmental structures and structural changes establish the wider contextual environment but do not trigger migration decisions per se. Socio-economic inequality, for instance, is predisposing but not directly triggering migration decisions.

More *proximate* drivers 'downscale' and localise broader macro-structural dispositions and are less abstract but closer to the immediate 'perception and decision spheres' of potential migrants. Macro-level structures and broader developments translate into more personally relevant factors at the meso and micro level while the ultimately *triggering* factors of migration are the actual reasons why people decide to migrate. This can be joblessness or job offers, marriage, exposure to threat or persecution, loss of assets, etc.

Beyond the degree of *immediacy*, migration drivers and driver configurations can further be characterised and categorised by their temporality, elasticity, selectivity, and geography. *Temporality* is referring to the less-than-permanent or transitory character of a specific driver or the volatility of a driver environment. For instance, demographic transitions or adaptations of socio-cultural norms are usually rather slow-changing and therefore relatively non-volatile structural drivers while natural disasters, or a coup d'état, are phenomena resulting in rapidly changing driver environments ("shocks"). Yet, slow-changing drivers are not less important than actual 'triggers' of migration in forecasting migration. They establish, through a cumulative process over time, the migration-conducive environment which may make migration in the future more likely, in particular when a sudden shock event, such as political turmoil, occurs. Slow-changing drivers are particularly important in long-term forecasting as they establish resilient processes of limited volatility.

Geography refers both to the locus and scope of a migration driver. The geographical *scope* of a structural driver can be anything between local and global, while the *locus* of a migration driver refers to the geographical contexts of both the origin and potential destinations but also in-between places of 'transit' where drivers are impacting on migration decision-making. The *selectivity* refers to the fact that broader social, economic or political transformations do not homogeneously affect all societal groups in the same way and to the same extent. Business cycles, for instance, affect societal groups in very different ways and to a varying degree depending on multiple intersections of age, gender, ethnicity, social status, or profession. Consequently, when a set of migration drivers constitute a specific driver environment, it conditions the time-space dependent *migration probability* of a homogenous group of people uniquely and uniformly exposed to a specific driver environment. The four key dimensions of migration driver functionality are summarised in Table 1.

Table 1. Functionality dimensions of migration drivers

Driver functionality	**...affecting migration probability**
Immediacy	predisposing *vs* mediating *vs* proximate drivers
Geography	location-specific *vs* global drivers
Temporality	slow-changing *vs* sudden-onset drivers
Selectivity	population-wide *vs* group-specific drivers

Source: Own elaboration

No single reason or driver, but complex multidimensional combinations of economic, political, social, and other developments and events may dynamically and heterogeneously influence (perceptions of) migration opportunities of alternative groups of people. In other words, time-space-dependent configurations of *complex driver environments* define people's willingness *and* ability to migrate (Czaika and Reinprecht 2020). It is also the accumulation of reasons (drivers) over time that lead to certain 'tipping point' situations at which migration of larger populations is triggered. For instance, many Syrians stayed in their home towns years into the civil war and only fled to neighbouring countries once their economic basis of subsistence eroded to an extent where staying was no longer a viable option.

Most empirical studies analysing migration drivers look at a very limited number of driving factors – on average only 2.5 drivers per study (Czaika and Reinprecht 2020). Only very few studies explore more complex combinations and interactions of migration drivers in shaping migration processes such as the interaction between migration policy and non-political factors (cf. Czaika and De Haas 2017; Düvell 2018; Van Hear et al. 2018) or between economic and security-related drivers (Crawley and Skleparis 2018). Interaction effects occur when the effect of one driver depends on the presence and intensity of another factor driving migration. That is, interaction effects indicate that other factors influence the (causal) relationship between a driver and migration outcome. Interaction affects the complexity of context-specific configurations of multiple drivers of migration. Besides, some combinations of drivers interact while other do not. The degree of mutual interconnectedness between drivers defines the overall stability (or resilience) of a migration driver environment.

Ragin (1987) explains that causal complexity as a phenomenon emerges from the intersection of relevant pre-conditions. In the absence of a necessary or sufficient 'ingredient', the phenomenon – e.g. migration – does not occur. In this set-theoretic logic, certain drivers can also be '*INUS*', that is *Insufficient*, but *Necessary* parts of an *Unnecessary* but *Sufficient* driver configuration (Mackie 1965; see also Czaika and Godin 2019). Identifying *relevant* migration driver configurations and assessing the ways and extent by which drivers intersect and interact must be fundamental elements of advanced migration forecasting analyses. At the same time, Ockham's razor principles apply: forecasting drivers bears its own uncertainty, which subsequently gets propagated to migration predictions (Bijak 2010). Hence, including too many proxy variables for drivers in forecasting models risks overfitting and actually reducing the predictability, instead of increasing it, whilst potentially providing the migration modellers and decision makers a false sense of accuracy resulting from the inclusion of many possible variables and their interactions.

3. Uncertainty in forecasting migration: a typology

Anticipatory knowledge on the characteristics of future migration can be convoluted by various types of uncertainty making the exact prediction of the number, form, duration, composition, direction, and locations of origin, transit and destination of future migrants an impossible endeavour. A realistic ambition for forecasting a multifaceted phenomenon like migration is not to predict the exact number of a specific group of people migrating in a specific period in a specific way between specific locations, but to estimate the *likelihood* – or, more broadly, a full *probability distribution* describing the future population movements. The analytical scope and degree of aggregation are hereby decisive elements affecting the level of accuracy of migration estimates.

Pooling certain populations, as long as they exhibit heterogenous migration behaviours and patterns, can reduce the uncertainty of estimation and prediction of the likelihood of migration. The law of large numbers allows more accurate predictions of future migration of large and highly aggregated populations than of largely disaggregated sub-populations. For instance, the relative margin of error in predicting total immigration of third country nationals into the European Union until 2025, is smaller than when predicting immigration of a specific nationality, occupation or a European destination in the same time period, with increasing relative margins of errors for time periods in the longer term. Similarly, aggregating over time can lead to less error overall because looking at pooled periods of several years rather than individual years has the advantage that annual errors would cancel out to some degree. when there is fluctuation on either side of linear forecasts.

Building on the overview of the complexity of migration drivers, presented in Section 2, in this section we elaborate on the different facets of the resulting fundamental uncertainty that are limiting the accuracy of future migration estimates – both epistemic and aleatory. These individual sources of uncertainty which are summarised in Table 2 across four domains: migration drivers, data, analytical methods, and individual decisions.

Table 2. Typology of uncertainty in the study of migration futures: A summary

Domain	**Epistemic uncertainty**	**Aleatory uncertainty**
Drivers	Conceptualisation: driver complexes	Systemic shocks in key driver configurations
	Future development of drivers	
Data	Measurement: migration and drivers	Shocks in methodological advancements and data availability
Methods	Analysis: model uncertainty	
Decisions	Decisions under uncertainty	Unpredictable behaviour

Source: Own elaboration

These key uncertainty *types* apply to all levels of aggregation and therefore affect estimates independently of the actual size of the population for which the likelihood of migration is assessed. The successive parts discuss the different types of uncertainty in more detail, following the key conceptual distinction between epistemic and aleatory aspects of the unknown migration futures.

3.1 Epistemic uncertainty

3.1.1 Conceptual uncertainty of migration driver complexes

What drives migration decisions on the one hand and larger population movements on the other? In the absence of a dominating analytical paradigm and a 'grand theory' of migration (Arango 2002), which arguably would not be very useful anyway, migration scholars refer to and attempt to triangulate a multitude of concepts, theories and epistemological approaches for understanding and explaining individual migration decisions and aggregate populations movements (see e.g. Massey et al. 1993, 1999, or Brettell and Hollifield 2014, for extensive but not exhaustive overviews of migration theories). Theoretically-informed identification and empirical assessment of the relevance of certain drivers of past migration is relatively well-established and largely feasible. Whether migration theories also have enough explanatory and predictive power to assess the multiple features of future migration however is less clear – and there are good reasons to be skeptical in that respect (Arango 2002).

As argued in Section 2, migration drivers are multifold and mutually interacting in often unexpected and counterintuitive ways, which is characteristic for all complex systems (e.g. Flood and Carson 1998). Based on a systematic review of the migration driver literature, Czaika and Reinprecht (2020) identify at least 24 different categories of migration drivers, which may all mutually interact in multifold ways. Accepting this number of distinct driving factors requires that a migration analyst needs to understand the relative importance of each driver in affecting individual migration and aggregate population movements. Assuming a binary measurement of all driving factors (e.g. high vs. low unemployment, restrictive vs. liberal border policy etc.), i.e. ignoring the continuous character of most migration drivers, we can imagine 2^{24} = 16,777,216 possible driver configurations or *driver environments* with each *ceteris paribus* potentially leading to slightly different likelihoods of migration, and ultimately, migration outcomes.

The complexity increases even more when we accept the fact that single drivers do not only operate independently but also in interaction with each other. For instance, the effects of immigration restrictions on the scale and composition of migration very much depends on the economic situation in both countries of origin and destination, but also in other relevant countries, such as transit ones. But moreover, if immigration policies are restrictive and economic prospects are dim, the existence of a mediating and reinforcing migrant network or a well-established 'culture of migration' may nevertheless lead to a perpetuation of immigration (Czaika and de Haas 2017, for a specific example related to the culture of migration to the US present in Mexico, see e.g. Kandel and Massey 2002).

What does configurational driver complexity and epistemic uncertainty imply for forecasting? Reducing driver uncertainty in migration modelling by including an almost infinite number of possible driving factors is certainly not a solution. In line with the Ockham's razor argument, a pragmatic solution is to identify a manageable number of driving factors selected according to their relative importance as either predisposing or proximate driver of migration. Assessing their relative importance in terms of a 'driver hierarchy' is thereby an analytical approach that involves both theoretical knowledge, as well as contextual and empirical assessment. Migration elasticity in terms of the migratory responsiveness of certain societal groups to changes in the driver environment can be an indicator for selecting relevant drivers in this process.

3.1.2 Uncertainty about future development of drivers

Migration drivers are generally not static but change dynamically. While some structural drivers can be rapidly changing (e.g. economic fluctuations, government policy), other may change only gradually (e.g. demographic transitions, cultural norms, climate conditions). Consequently, uncertainty regarding the future development of migration drivers and driver environments is a central element of driver uncertainty, which needs to be taken into account in migration predictions.

Even if epistemic knowledge regarding certain drivers and their (past) functionalities is reliable and truthful, and if uncertainty about their measurement can be assessed and controlled for, the state of single drivers and driver configurations are likely to change in the future, with the likelihood of change usually increasing with time horizon. However, considering knowledge and empirical evidence about 'temporal regularities' and the nature of dynamic changes of certain drivers, analysts can identify and separate slow-changing from fast-changing drivers, as well as stationary from non-stationary processes (Bijak and Wiśniowski 2010). For instance, by the nature of their temporal dynamics, future prediction of slow-changing transitions from high to low fertility rates or from an agriculture-based to an industry-based economy have smaller margins of error than more cyclical economic fluctuations or electoral outcomes and implications for political stability.

3.1.3 Uncertain data and measurement of migration and its drivers

The uncertainty of migration starts right with conceptualizing and formally defining what is a migration event, and who is a migrant. Unlike in the case of other key demographic processes, which involve events that are uniquely defined (birth, death, marriage, divorce), migration needs to be conceptualized with reference to its two key dimensions: *space* (how far a person needs to move to become an internal or international migrant and what administrative boundaries they need to cross), and *time* (for how long). Both these dimensions require putting arbitrary boundaries to delineate migration from other types of mobility. This leads to other issues, such as the Modifiable Area Unit Problem in space or its equivalent in time, where the number of migrants observed depends on the spatial or temporal granularity of observation (see Rees and Lomax 2020, as well as Nowok and Willekens 2011 for the temporal discussion). This ambiguity is reflected in a range of definitions of migration prevalent across different countries, despite attempts at harmonizing them, such as the UN (1998) standards, or the EU regulation 862/2007 on migration and asylum statistics[8].

The definitional patchwork is often a result of different national or supra-national legislative frameworks, which often vary between the countries with regard to the purpose and mechanism of migration data collection. In addition, migration statistics can be reported by the country of previous or next residence of migrants, their country of birth, citizenship, or legal status, such as visa, formal residence, or asylum (see Lanzieri 2019b). At the same time, it has to be borne in mind that migration is a *process*, although what gets measured is typically individual *events* or *transitions* from one place to another. This is a great simplification, which can help with modelling, but removes the measurement even further from the messy social reality with its long-distance cross-border commuting and circular migration, transnational ties, and so on.

---

[8] Regulation (EC) No 862/2007 of the European Parliament and of the Council of 11 July 2007 on Community statistics on migration and international protection; OJ L 199, 31.7.2007, p. 23–29, with further amendments.

One crucial element of epistemic migration uncertainty is related to its measurement. To start with, migration statistics are social constructs, reflecting political priorities prevalent at the time of their collection (Bijak and Koryś 2009). This results in the presence of different data collection tools and mechanisms, as well as their different focus, which does not have to be on measuring migration as such, but on other purposes of data collection, such as law enforcement or facilitating some administrative processes (e.g. tax, social security, or medical records). Typically, migration data sources include censuses, surveys and administrative data, currently also making a foray – albeit still to a limited extent – to the world of 'big data' and digital traces, such as mobile phones or social media (Cesare et al. 2018). Different sources of data have different problems with respect to the timeliness of data production, and a range of other quality aspects: coverage of specific populations and national groups, the presence of various biases and possible under- or overcount, and accuracy of the measurement instruments (Raymer et al. 2013).

All these features may require a recourse to formal modelling as a way of combining information from different data sources, bringing them to a common denominator, and describing the resulting uncertainty in a coherent manner (see e.g. the International Model of European Migration project IMEM of Raymer et al. 2013, the continuation of which is envisaged within the current programme of work[9]). Given the epistemic nature of this type of migration uncertainty, it can be at least described, but hopefully also to some extent reduced through improvements in data collection and analytical tools. Such data collection advances can including for example the economical option of 'mainstreaming' of migration and asylum by including related questions and variables in more general data sources (Knauth 2011), as well as more migration-centred approach, with surveys specifically dedicated to examining different features of migration.

Likewise, the data on migration drivers and correlates are also fraught with measurement and comparability issues. In addition, attempting to explain or predict migration – highly uncertain itself – with the help of imperfect proxies for drivers (e.g. GDP ratios for income differentials) adds another layer of uncertainty, which should be propagated into the final results of the analysis together with migration measurement errors. One problem here is that in applied work, the uncertainty from exogenous variables gets lost, leading to unreliable conclusions (e.g. Dustmann et al. 2003), or when it gets included in a simple way, the resulting uncertainty is overpowering, rendering the results useless (Bijak 2010).

3.1.4 Formal analysis: Model uncertainty

Every formal model is an approximation of the reality under study. Models include errors of different kinds: in model specification, model parameters, parametric uncertainty (error term), model discrepancy, code uncertainty, and more (Kennedy and O'Hagan 2001). In addition, if estimation, forecasting or scenario setting involves expert knowledge, the breadth of subjective expert opinion adds an additional source of uncertainty. This uncertainty can be formally elicited and described (see O'Hagan et al. 2006), by which it becomes epistemic, reflecting the current state of expert knowledge on the migration processes under study.

[9] QuantMig work package 6, see http://quantmig.eu for details.

In forecasting – as opposed to using models to explain past trends and behaviours – the focus is clearly on predictive 'out-of-sample' performance, which varies with the time horizon of prediction. This performance can be measured by using a range of metrics, from errors (absolute, relative, and so on), to coverage calibration measures, and other scoring rules, such as Brier scores or similar (Gneiting and Raftery 2007). In the effort to reduce errors, over the past decades, the models of social processes, including migration, became increasingly sophisticated. Still, this does not mean that their uncertainty can be reduced (see e.g. the complexity-simplicity debate in Ahlburg 1995, and Smith 1997; as well as Keilman 2008) – rather, than at best it can be described more realistically.

In general terms, the uncertainty of a model can be assessed based on its past performance, with *ex-ante* (theoretical, model-based) and *ex-post* (empirically observed) error distributions ideally aligned, and the overall objective to minimize the latter. It is worth remembering that the goal is not to minimize uncertainty or pretend it does not exist, which would lead to an 'illusion of control' (see e.g. Makridakis and Taleb 2009) – but to describe it more realistically, bearing in mind the different sources, so that it better severs prudent policy and decision making.

An additional problem in modelling migration drivers concerns endogeneity: not all variables that are used as proxies for exogenous drivers in models are indeed exogenous. One clear example is population (e.g. Cappelen et al. 2015), which changes deterministically through migration as well as being an important migration driver. Alternative approaches, based on statistical models, which do not rely on drivers, can only offer approximate solutions (Bijak 2010; Azose and Raftery 2015). Such approximations, for example those based on time series techniques, are useful as they bypass the problem of theoretical complexity and rely on the persistence and self-perpetuation of migration (Bijak 2010). These features, broadly related to the notion of cumulative causation (Myrdal 1957), are well-acknowledged elements of existing explanations of migration processes (see e.g. Massey et al. 1993), and can be described by autoregressive models. On the other hand, the uncertainty related to driver measurement is largely epistemic, and potentially reducible with the development of new data sources, analytical tools and methods, for example in the form of more highly structured models, combining information from as many sources as possible.

3.1.5 Migration decisions under uncertainty

At the individual level, migration decisions are taken in the context of personal uncertainty and risk. Information about individual prospects and opportunities is incomplete, and whether migration turns out as the 'right' decision depends very much on circumstances that are *ex ante* unknown and *ex post* not fully under control of the migrant (Czaika 2015). Migration decisions are mostly based on *limited* information and uncertainty about migratory options, possible outcomes and personal preferences. In addition, the non-static nature or 'endogeneity' of personal preferences and perceptions about driver environments involving opportunities and constraints has hardly been explored and considered in migration analyses.

Risk and uncertainty in the context of migration decision-making has been addressed by several authors (e.g. Williams and Baláž 2012), some suggesting that migrants tend to be rather risk-loving (e.g. Sahota 1968) than risk-averse, as assumed in neoclassical expected utility theory. Migrants are supposed to consider risk and uncertainty inter-temporally, e.g. by trading medium-level risks for immediate higher risks, but subsequent lower risks (Katz and Stark 1986). However, all these propositions refer to the expected utility theory and the standard neoclassical model with its mostly unrealistic assumptions about fully rational decision-making. The theory of revealed preferences also suggests that migrants choose a destination because it is preferred to any other destination if migration to that other destination was both possible and affordable. However, real preferences and revealed preferences (e.g. for alternative migration destinations) are not necessarily identical.

What makes the anticipation of migrant behaviour even more uncertain than expected utility theory would suggest, is basically the fact potential migrants are 'bounded rational' decision-makers, implying that human rationality is limited and constrained by the situation, incomplete information and cognitive limitations (Simon 1983). Furthermore, preferences and risk attitudes are known to be non-static but dynamic, endogenous and reference dependent. This proposition of prospect theory (Kahneman and Tversky 1979, Kőszegi and Rabin 2007) suggests that potential migrants may tend to change risk attitudes by being risk-averse when facing positive migration prospects (such as increased income or job opportunities), and risk-friendly when expecting negative prospects (such as joblessness or persecution; for a discussion, see also Czaika 2015). Thus, migrants adjust their risk attitudes depending on their individual reference point and the type of information they receive about real or expected outcomes. Individual reference points are space- and time-dependent and group-specific so that a formulation of a migration probability function would require very specific information on group-specific reference points for assessing migration-related prospects.

A key aspect in an individual migration decision-making processes is the search for migratory options. Obviously, neither all possible migration alternatives nor all consequences can be individually known, but normally a 'manageable set of options' is identified, investigated, and evaluated. Search for information is complex and involves collection and processing of information about possible migration destinations and corresponding opportunities including information on possible entry routes, job and livelihood opportunities, but also risk and uncertainty related to the migration journeys and outcomes. While neoclassical random utility models (RUM) often unrealistically assume migrants' full information e.g. on all possible migration options, searching for this kind of information in fact places high burdens on individuals (Gigerenzer et al. 1999).

Due to the complexity and the amount of information, potential migrants apply instead some simple decision rules or heuristics in the migration decision-making process (Tversky and Kahneman 1974; Czaika and Reinprecht, forthcoming). Heuristics are "methods for arriving at satisfactory solutions with modest amounts of computation", which are often used unconsciously or automatically (Simon 1990, p 11). Heuristics are strategies that aim to make decisions "more quickly, frugally, and/or accurately than more complex methods" by ignoring part of the information (Gigerenzer and Gaissmaier 2011, p 454).[10] Expected utility theory and respective assumptions about preferences and probabilities struggles to explain 'behavioural anomalies' (Kahneman et al. 1991) such as endowment effects, loss aversion, status quo bias, or usage of other heuristics in migration decision-making (Czaika and Reinprecht, forthcoming).

[10] See, for instance, Koser (1997) on the role of information and uncertainty in the context of return migration.

3.2 Aleatory uncertainty

3.2.1 Unpredictable systemic shocks in key driver configurations

Aleatory uncertainty refers to the inherent uncertainty due to the probabilistic variability or other types of randomness, such as chaotic behaviour. This type of uncertainty is non-reducible, that is, there will always be unpredictable variation in single migration drivers and broader driver configurations and the way these are shaping migration processes. For instance, the unpredictable nature of sudden shocks to social systems such as a natural disaster, a terrorist attack, or an outbreak of violence, and its uncertain implications on migration creates a key category of aleatory uncertainty which affects forecasts of future migration.

So-called 'black swan events' (Taleb 2007) are unpredictable and go beyond what is normally expected of a situation and has potentially significant consequences for society and (some or all) its members. Black swan events are characterized by their extreme rarity and their severe impact. Rare events or shocks suddenly change the driver environment that may affect migration intentions of larger groups of people in a positive or negative way. In contrast to 'normal' aleatory uncertainty, standard tools for evaluating probability and prediction such as models based on thin-tailed normal distributions do not apply. Such tools depend on large numbers of observations and sample sizes (here, long data series) that are – by definition – not available for rare events.

Even though statistical techniques exist to deal with some aspects of this unpredictability, for example by assuming heavy-tailed distributions or explicit modelling of extreme values (Coles 2001), they only offer partial solution to the related challenges of migration forecasting. Even if we know that rare events happen, we cannot predict their exact timing, location, magnitude or impact. At the same time, there is a residual part of unpredictability, which evades statistical modelling, which is partially related to the inadequacy of models of any complex systems (Frigg et al. 2014)[11]. In either case, as with earthquakes or similar rare events, we can still prepare contingency plans to deal with the occurrence of rare, yet high-impact events, shifting the focus from *prediction* to *response*.

In rare-event situations, the cause (or rather, the causal driver configuration) of migration is unpredictable, and so are the migratory responses: partly due to the rarity of the events in question, and partly to the fact that complex systems are difficult to model. At the same time, 'what-if' assessments and preparatory actions of various actors and stakeholders (including potential migrants themselves) may affect and mediate the actual impact of a rare event on migration. Again, this is where the boundary between aleatory and epistemic uncertainty may become a bit blurred: even though the exact timing, location or magnitude of black swan events may not be predictable, the underlying processes operating on longer time scales (such as climate change) make the *occurrence* of rare events at least probable in the long term[12]. This helps assess at least the impact of such events on migration, if not their likelihood. The related risk assessment process needs to be detailed and specific to a particular situation or migration policy challenge. This is similar to the way

[11] Historically, this has led to the evolution of systems science and complexity science as separate fields of enquiry, offering a range of model-based approaches to explore the dynamics of complex systems (see e.g. Flood and Carson 1998).

[12] For instance, a once-in-a-decade event can be expected to happen on average three times between 2020 and 2050. If such events result from underlying processes that operate on longer timescales, such as the climate change, the *likelihood* of their occurrence, as well as their *impact*, or their *volatility*, may become more predictable. (With thanks to Frans Willekens.)

in which reinsurance companies price the risk they underwrite, which is very much tailor-made for individual circumstances (Clark 2014). This perspective on formal treatment of rare events is one important gap in the current literature on migration forecasting.

3.2.2 Unpredictable shocks in methodological advancements and data availability

Conceptual and methodological advances in explaining and predicting migration have so far been largely incremental. There is no scope for – and likely never can be – a single 'grand theory of migration', which would need to integrate too many different perspectives to be operationally feasible (Arango 2002), while still leaving out the important unpredictable aspects. Hence, despite important advances in the conceptualization of driver environments and complexes, as discussed above, some of the associated uncertainty remains irreducible given the sheer complexity of migration processes.

In terms of data and methods, in the recent years important progress has been made: the availability of new, more timely data sources (traces form mobile phones, social media, to mention but a few) offers additional, much more agile opportunities for policy and research, such as in the near-term prediction or 'nowcasting' or providing early warnings. Still, given the volatility of such data sources, they are unlikely to offer much help in the long-run scenario planning. Similar arguments hold for methodological advances: there is a lot of potential in having more fine-grained methods, using advanced statistical modelling, artificial intelligence or data mining techniques, as well as exploring causal models of migration mechanisms (Willekens 2018).

Nevertheless, even if there is a 'step-change' epistemological shock in terms of more precise measurement and short-term forecasting methods based on pattern detection and more accurate separation of the migration and driver signal from the surrounding noise, capability for any longer-range prediction is a different story. The fundamental inherent uncertainty of the future – firmly aleatory in nature – remains unchanged, given the fundamentally indeterministic character of our world – both physical and social – and the free will of people as key decision-making actors (see e.g. Popper 1982[13]). Some of that indeterminism can be expressed and modelled probabilistically, while other aspects, linked to the more unstable processes, require extra caution when attempting to model them and draw conclusions on that basis (e.g. Frigg et al. 2014). In addition, in situations where the model error dominates over more regular, statistical randomness (*idem*), exploration of the robustness of alternative modelling approaches, such as ensemble methods (Gneiting and Raftery 2005) or Bayesian model averaging (Hoeting et al. 1999), is worthy of consideration.

3.2.3 Unpredictable behaviour

Besides the shocks to the broader migration systems that happen at the macro level and affect whole societies or specific groups (such as economic conditions, war and persecution, and so on), an important aspect of aleatory uncertainty is related to the individual-level decision making. The advances in behavioural science notwithstanding, people have their own agency as decision makers, which is again linked to the free will and choice (Erdal and Oeppen 2018). As such, it is worth

[13] See especially Popper's (1982) Section 20 on "Historical Prediction and the Growth of Knowledge" and the Addendum: "Indeterminism is Not Enough: An Afterword".

repeating that these are ultimately the migrants who make decisions about migrating – not the policy makers – and this is one of the key reasons underpinning the migration policy failures, according to Castles (2004). This results in unintended consequences of policies, stemming from interactions between the decisions of multiple agents with different objectives (Czaika and de Haas 2013).

Besides, there is an important element in what in macroeconomics became known as the Lucas critique (Lucas 1976): forecasts self-invalidate, because policy makers and other actors act in response to predictions. There is scope for methodological work in building micro-foundations into the models of migration (e.g. through agent-based models; see Klabunde and Willekens 2016, for a recent overview). Still, despite some regularities seen in the aggregate statistics, unpredictability of the human behaviour remains largely irreducible, especially at the individual level. Hence, the best we can hope for in the context of aleatory uncertainty is a better approximation of the underlying macro-level migration processes, for example by using simulation models, or – in turn – their approximations based on statistical meta-models (emulators, see e.g. Pietzch et al. 2020 for a recent review). In the case of such complex models and systems as migration, we cannot expect to eliminate or vastly reduce the associated uncertainty, which also to a large extent aleatory[14].

[14] For a broader discussion about consequences of model error for prediction, see e.g. Frigg et al. (2014).

4. Future trends: Possibilities and limits of forecasts

Given what is known about the migration processes and their underpinning factors and drivers, what are then the perspectives for predicting different types of flows, and for assessing their uncertainty by using a common analytical framework? Even if predicting migration in the strict sense is impossible, and in the qualitative sense remains at least difficult, there are things that can be said today about the different possible migration futures. To that end, as suggested in the typology proposed in this paper, being able to distinguish between the epistemic uncertainty, depicting our limited knowledge about the migration reality; and an intrinsic aleatory one, related to the unpredictability of human and social behaviour, is of key importance.

Migration processes are characterized by both types of uncertainty to a large extent, which contributes to their barely predictable nature. Still, different types of flows exhibit a different mix of epistemic and aleatory aspects. For example, migration to study has important epistemic elements, being largely driven by the number of places on courses that can be offered to foreign students and having considerable inertia, whereas uncertainty of asylum-related migration induced by armed conflict will be more aleatory, due to the unpredictable nature of its driver environments.

As an illustration, for the current COVID-19 pandemic and its possible impact on migration, both aspects of uncertainty are strongly at play. Here, while the epistemic aspects enable linking changes in future migration to reductions in travel, mobility and economic activity, at least in the shorter term, the more aleatory ones concern other aspects, such as viral mutations, as well as possible responses to the pandemic. The latter may include for example by permanent shift to predominantly online work and study, or the development of efficient vaccinations within in a specific time horizon, which at the time of writing still remain largely uncertain.

At the micro level, migration decisions are taken in the context of personal needs and preferences, perceived opportunities, and uncertainty of both. The changing nature and consciousness about personal needs, preferences, and aspirations, e.g. regarding ideas of 'the good life', leaves the *prediction* of individual behavioural responses for realising it an almost impossible endeavour. On the other hand, there exist visible epistemic regularities, such as those linking migration to the life course (Courgeau 1985; De Jong and De Valk 2020), with resulting stable age profiles of migration (Rogers and Castro 1981). As a result, the degree of epistemic and aleatory uncertainty of migration decision-making at the individual level is higher than uncertainty regarding behavioural actions of groups or populations. Uncertainty can be better handled at higher levels of aggregation, where we can gauge more epistemic aspects by using statistical methods to better understand the nature of the processes and trends, than at more fine-grained levels. Consequently, disaggregation of global migration by type of flows, period, geographical entity, or various other group-specific intersections, increases the degree of uncertainty for forecasting, unless the sub-groups are more homogenous and share more similarities within than between groups, where disaggregation can help[15].

[15] See, for example, the disaggregation by migration types (e.g. labour, study, family, asylum) proposed by de Beer (2008) to improve predictability, although the wider argument is more general, and links with Keyfitz's (1985) "search for constancies" in the analysis of trends to be forecast. (With thanks to Nico Keilman.)

The different aspects of epistemic uncertainty clearly point to the priority areas for the future policy and research agenda. Academic and applied work can therefore focus on some of the following questions and topics:

- Clarifying the ambiguity of concepts (who is a migrant and what is a migration event?) and designing a flexible, dynamic framework for defining migration, which is of foundational and critical importance.
- Dealing with the known shortcomings of different theories of migration and limited connections between them, and further exploring multidimensional driver complexes and environments (why do people migrate and what decision processes are involved?)
- Addressing widespread problems with the collection, quality and other measurement aspects of migration data (how many people move, and for what purposes?). Here, one promising approach is to combine and harmonize information from different sources, while being explicit about the measurement uncertainty (Raymer et al. 2013).
- Pursuing new, promising avenues in terms of methodological innovation for migration modelling, explanation, prediction and scenario-setting including causal models, model ensembles, simulations and other innovative approaches (Willekens 2018).
- Further exploring the various regularities of migration in terms of their spatio-temporal patterns (first noted by Ravenstein 1885, see also Rees and Lomax 2020), age schedules and links with the life course (Courgeau 1885; Rogers and Castro 1981), persistence and self-perpetuation of existing flows (Massey et al. 1993), or the relative stability of the remittances sent by migrants (Neagu and Schiff 2009; World Bank 2019)[16].
- In addition, testing the boundaries of the epistemic uncertainty, to make sure that the classification into the epistemic (knowable) and aleatory (unknowable) types remains correct and relevant to the research questions and policy challenges at hand.

To effectively deal with the epistemic uncertainty, a crucial and necessary starting point is self-awareness: a clear acknowledgement of the limitations of knowledge and imperfections of the sources of data and other information used in the analysis, including expert judgement. An initial assessment of the limits of current knowledge can help improve the understanding of the migration processes, patterns and drivers, whilst remaining cognizant about the limitations of the inferences that can be drawn from the available information. A clear research agenda related to migration scenario setting for the future includes pushing the boundaries of the epistemic uncertainty and shedding more light on the knowable aspects, such as the observed regularities, links with other aspects of the life course, and so on.

On the other hand, there remains the aleatory uncertainty, related to the multiplicity of drivers, often interacting with one another and operating at a range of time scales, as well as to the human agency of people making decisions, also under the conditions of uncertainty. It means that migration can be very volatile, and respond very rapidly to economic or political shocks, as discussed above. The various facets of aleatory uncertainty cannot be reduced by improving our knowledge about the migration processes and the context in which they occur – but still, this type of uncertainty can be

[16] This strand of work could additionally benefit from links with more qualitative studies, e.g. on lifecycle patterns, intergenerational relationships, remittance behaviour, and other. (With thanks to Marta Bivand Erdal)

more effectively managed than it is the case today. Ignoring it can backfire in the face of rapidly changing events due to the 'illusion of control' which can leave the decision makers unprepared (Makridakis and Taleb 2009).

In the very short-term horizon, there is additionally scope for developing early warning systems, which would help in the operational management of potentially rapidly changing migration flows (see e.g. Bijak et al. 2017)[17]. For the short- and mid-term planning, some numerical input is typically required, which should be coupled with some (ideally probabilistic) assessment of error. Expressing uncertainty about future migration in the language of probabilities conveys an additional message about the current limits of our knowledge. In addition, probabilistic statements help explicitly measure by how much we expect *ex ante* to be wrong in our predictions, and how often we can expect errors of certain magnitudes. In the longer run, for example years or decades ahead, future trends based on different scenario techniques, both narrative and simulation-based, can help, as long as they are bold and imaginative enough to cover at least some of the aleatory uncertainty.

In all instances, the estimates, probability distributions and scenarios offer but one input for the decision process – the other being explicit policy preferences, for example readiness to prepare for different migration contingencies and hedge against various negative outcomes, and to commit adequate resources for that purpose. This is an element of political and social choice, which remains outside of the research-led advice on the possible migration futures. Still, it is the responsibility of the forecasters and scenario-makers to convey the message about uncertainty to the final users: we cannot know how many migrants from different groups will arrive or leave a country in the future, but we can put some reasonable bounds on the numbers, together with a measure of their accuracy.

From the point of view of the end-users of migration forecasts and scenarios, there is also scope for a wide range of tools and approaches for aiding the relevant decisions and testing the various possibilities, depending on the purpose and time horizon and the predictability of flows. The available options for shorter-range operational and planning decisions and for those related to the better-predictable flows, made chiefly under epistemic uncertainty include statistical decision analysis based on probabilistic forecasts and user-specific loss functions (Bijak 2010). The decision analysis would then support a formal risk-benefit assessment of various policy options, related to more operational questions such as how many new school places or hospital beds are needed given the expected migration, or where to focus how much humanitarian effort for asylum flows.

On the other hand, the longer-term strategic thinking in the aleatory context can benefit from interactive scenario-based explorations of different 'what-if' assumptions (*idem*), which is crucial for stress-testing of policies and making contingency plans, aiming at building resilience of the systems involved under potentially very volatile conditions. Any such policy instruments should ideally be developed in close collaboration between policy makers, stakeholders from civil society

[17] As argued before, with processes that are sufficiently slow and operate on longer timescales, such as population or climate change, it may be even possible to predict events for more distant horizons, such as years rather than days, as long as these horizons are still relatively short when compared to the secular nature of the underlying processes. This also extends the definition of an early warning system to a model that signals a possibility of an event (e.g. migration) occurring on a given timescale, based on changes in the underlying or associated processes, which can operate on different timescales. (Thanks to Frans Willekens for this argument.)

organizations, and academia. At the same time, as alluded to in the Introduction, it is worth bearing in mind that migration, even if its various impacts need accommodating, can bring about advantages as well as challenges. An honest discussion about the uncertainty and possible impacts of migration processes can additionally help enhance the quality of the public debate surrounding migration, by making the available public choice options and their possible consequences open and transparent.

In general, the appeal of both probabilistic forecasts and variant 'what-if' scenarios is that they underline the non-deterministic nature of migration, and that its futures are not restricted to a single possible path. Such tools provide a natural way to communicate plausible ranges of future trends, with probabilistic approaches additionally offering a measurement tool in the form of probability distributions or error terms. In a small survey of Australian users of population forecasts, Wilson and Shalley (2019) found that 61% of respondents made use of information to provide guidance about forecast uncertainty, with the knowledge about uncertainty and prediction intervals increasing with the experience in using population forecasts (*idem*).

From the forecasters' side, there is also a challenge of trying to improve the accuracy and calibration of some predictions made under uncertainty. In their bestseller on 'superforecasters', Tetlock and Gardner (2015) argue that many forecasting exercises involve an element of skill. Nevertheless, this pertains chiefly to predictions made under epistemic uncertainty, and only in the short horizon of weeks or months, rather than years (*idem*), which are too near-term for strategic-level scenario planning, ideally managing the aleatory aspects as well[18]. An additional conceptual challenge here is the need to delineate the aleatory and epistemic forms of uncertainty more clearly, and to further work on reducing the latter, whilst acknowledging and being mindful of the former.

To help manage this dilemma, an important strand of ongoing work on the uncertainty for migration scenarios[19] concentrates on what can be done across a range of time horizons – from days or weeks in the case of early warnings, to decades for scenario-based planning. In particular, techniques like stationarity and cointegration analysis, dynamic stochastic general equilibrium (DSGE) models, early warnings and signal detection tools, and many other would need to be explored in the context of migration scenarios, as the previous attempts (Bijak 2010; Bijak et al. 2017) barely scratched the surface of the many possible methodological opportunities in that respect.

As mentioned before, the questions of rare events, their impact of migration, and potential for applying risk management tools in the long horizons, also require research attention. As estimates of uncertainty are themselves uncertain, research into higher-order uncertainty may also be of interest, but pragmatically, it is probably worth stopping at modelling the second-order errors as a 'good enough' approximation of the overall uncertainty assessment.

Finally, irrespective of the type of uncertainty associated with particular features of population flows – aleatory or epistemic – more work needs to be done on ways to manage and adapt to migration

[18] A very similar argument can be made regarding the usefulness of prediction markets in assessing future migration flows, which can work only in very short time horizons, which are closest to fulfilling the efficient market hypothesis conditions, with limited usefulness in a longer term. Besides, the track record and success rate of prediction markets even for well-publicised high-stakes political events remains mixed (for examples, see the discussion related to two recent referenda in the UK, in Wall et al. 2017 and Auld and Linton 2018).
[19] QuantMig work package 9, see http://quantmig.eu for details.

processes, and to anticipate and address their most significant impacts. Clearly, this is more pertinent for the aleatory types, where there are few other options but ensuring adequate resources, contingency plans, and enhancing preparedness. Still, reducing the epistemic uncertainty will likely take time and research effort, and the results are far from guaranteed. What is clear that migration exhibits a considerable 'uncertainty about uncertainty', especially insofar as any attempts at quantification are concerned. This strengthens the need for proper understanding of any formal models and approaches used for prediction, and their limitations[20]. In this case, increased focus on managing responses to changing migration under the conditions of uncertainty becomes a prudent policy solution, with a more universal appeal.

In parallel, research work done both at the high, conceptual level, as well as at the detailed, empirically relevant level of description, can help contribute to painting the future map of migration forecasting. After all, some of the big ideas, such as Wilbur Zelinsky's (1971) mobility transition theory continue to be relevant in the world of constant surprises, with the argument about exchangeability between migration and other forms of long-range human contact, communication and work, yet again gaining traction in the world preparing to emerge from the COVID-19 pandemic into a possibly new reality for human mobility.

[20] With thanks to Giampaolo Lanzieri; for a recent, more general discussion, see Saltelli et al. (2020).

References

Acostamadiedo E, Tjaden J, Sohst R, de Valk H and Groenewold G (2020) Future European Migration Scenarios until 2030: Plausible, Probable and Relevant? CrossMigration project report. Berlin: IOM GMDAC.

Ahlburg DA (1995) Simple versus complex models: Evaluation, accuracy and combining. *Mathematical Population Studies*, 5(3), 281–290.

Arango J (2002) Explaining migration: a critical view. *International Social Science Journal*, 52(165), 283–296.

Auld T and Linton O (2018) The behaviour of betting and currency markets on the night of the EU referendum. CEMMAP working paper CWP 01/18, available at https://www.ifs.org.uk/uploads/CWP011818.pdf (accessed on 14 June 2020).

Azose JJ, Raftery AE (2015) Bayesian Probabilistic Projection of International Migration. *Demography*, 52(5), 1627–1650.

Bijak J (2010) *Forecasting International Migration in Europe: A Bayesian View*. Dordrecht: Springer.

Bijak J, Forster JJ and Hilton J (2017) Quantitative assessment of asylum-related migration: A survey of methodology. Report for the European Asylum Support Office. Malta: EASO.

Bijak J, Disney G, Findlay AM, Forster JJ, Smith PWF and Wiśniowski A (2019) Assessing time series models for forecasting international migration: Lessons from the United Kingdom. *Journal of Forecasting*, 38(5), 470–487.

Bijak J and Koryś I (2009) Poland. In: H Fassmann, U Reeger, and W Sievers (eds.), *Statistics and reality: concepts and measurements of migration in Europe*. IMISCOE Reports. Amsterdam: AUP, 195–215.

Brettell CB and Hollifield JF (eds.) (2014) *Migration theory: Talking across disciplines*. London: Routledge.

Cappelen Å, Skjerpen T and Tønnessen M (2015) Forecasting Immigration in Official Population Projections Using an Econometric Model. *International Migration Review*, 49(4), 945–980.

Carling J and C Talleraas (2016) Root causes and drivers of migration, PRIO Paper, Oslo

Cesare N, Lee H, McCormick T, Spiro E and Zagheni E (2018) Promises and Pitfalls of Using Digital Traces for Demographic Research. *Demography*, 55(5), 1979–1999.

Chojnicki Z (1970) Prediction in Economic Geography, *Economic Geography*, 46(2: Suppl.), 213–222.

Clark DR (2014) Basics of Reinsurance Pricing. Actuarial Study Note. Mimeo; available via https://www.casact.org/library/studynotes/Clark_2014.pdf (accessed on 14.06.2020).

Coles S (2001) *An Introduction to Statistical Modeling of Extreme Values*. Berlin: Springer.

Courgeau D (1985) Interaction between Spatial Mobility, Family and Career Life Cycle: A French Survey. *European Sociological Review*, 1(2), 139–162.

Czaika M (2015) Migration and economic prospects. *Journal of Ethnic and Migration Studies*, 41(1), 58–82.

Czaika M and Godin M (2019) Qualitative Comparative Analysis for migration and development research, MIGNEX background paper, Oslo

Czaika M and de Haas H (2017) The effect of visas on migration processes. *International Migration Review*, 51(4), 893–926.

Czaika M and Reinprecht C (2020) Drivers of migration: A synthesis of knowledge, IMI Working Paper No. 163, University of Amsterdam

Czaika M and Reinprecht C (forthcoming) (Non-)Migration in the context of environmental stress, Danube University Krems.

de Beer J (2008) Forecasting international migration: time series projections vs argument-based forecasts. In: J Raymer and F Willekens (eds.), *Estimation of international migration in Europe: issues, models and assessment*. Chichester: John Wiley (pp. 283–306).

De Haas H, Vargas-Silva C and Vezzoli S (2010) Global Migration Futures: A conceptual and methodological framework for research and analysis. International Migration Institute, University of Oxford, Oxford

De Jong PW and de Valk HG (2020) Intra-European migration decisions and welfare systems: The missing life course link. *Journal of Ethnic and Migration Studies*, 46(9), 1773–1791.

Dustmann C, Casanova M, Fertig M, Preston I and Schmidt CM (2003) The impact of EU enlargement on migration flows. Home Office Report 25/03. Home Office, London.

Erdal MB and Oeppen C (2018) Forced to Leave? The Discursive and Analytical Significance of Describing Migration as Forced and Voluntary. *Journal of Ethnic and Migration Studies*, 44(6), 981–998.

Fitzgerald T (2018). Building Predictive Models to Forecast Population Trends in Ireland for 2017–2021. Master's thesis. National College of Ireland: Dublin.

Flood RL and Carson ER (1998) *Dealing with Complexity: An Introduction to the Theory and Application of Systems Science*. Second edition. New York: Springer.

Foresight (2011) Migration and Global Environmental Change: Future Challenges and Opportunities. Final Project Report. London: Government Office for Science.

Frigg R, Bradley S, Du H and Smith LA (2014) Laplace’s Demon and the Adventures of His Apprentices. *Philosophy of Science*, 81(1), 31–59.

Frontex (2016) Risk Analysis for 2016. Warsaw: Frontex.

Gigerenzer G and Gaissmaier W (2011) Heuristic decision making, *Annual Review of Psychology*, 62, 451–482.

Gigerenzer G, Todd PM and ABC Research Group (1999) *Simple Heuristics That Make Us Smart*. Oxford: Oxford University Press.

Gneiting T and Raftery AE (2005) Weather Forecasting with Ensemble Methods. *Science*, 310(5764), 248–249.

Gneiting T and Raftery AE (2005) Strictly Proper Scoring Rules, Prediction,and Estimation. *Journal of the American Statistical Association*, 102(477), 359–378.

Hempel CG (1962) Deductive-nomological vs. statistical explanation. In: *Scientific explanation, space, and time. Minnesota studies in the philosophy of science*, Vol. 3. Minneapolis: University of Minnesota Press (pp. 98–169).

Hoeting JA, Madigan D, Raftery AE and Volinsky CT (1999) Bayesian Model Averaging: A Tutorial. *Statistical Science*, 14(4), 382–417.

Holzer JZ (1959) *Prognoza demograficzna Polski na lata 1960–1975 wg województw* [*Demographic forecast of Poland for 1960–1975, by voivodships*]. PWG State Economic Publishers, Warsaw.

Jäger J, Frühmann J, Grünberger S and Vág A (2009) Environmental Change and Forced Migration Scenarios. Synthesis Report. EACH-FOR project.

Kahneman D, Knetsch JL Thaler RH (1991) Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias, *Journal of Economic Perspectives* 5(1), 193–206.

Kahneman D and Tversky A (1979) Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47 (2), 263–291

Kahneman D and Tversky A (1982) The Psychology of Preferences. *Scientific American,* 256 (1), 160–173.

Kandel W and Massey DS (2002) The Culture of Mexican Migration: A Theoretical and Empirical Analysis. *Social Forces*, 80(3), 981–1004.

Katz E and Stark O (1986) Labor Migration and Risk Aversion in Less Developed Countries, *Journal of Labor Economics* 4 (1), 134–149.

Kennedy MC and O'Hagan A (2001) Bayesian calibration of computer models. *Journal of the Royal Statistical Society B*, 63(3), 425–464.

Keilman N (2008) European demographic forecasts have not become more accurate over the past 25 years. *Population and Development Review*, 34(1), 137–153.

Keilman N and Pham DQ (2004) Empirical errors and predicted errors in fertility, mortality and migration forecasts in the European Economic Area. Discussion Papers No. 386. Oslo: Statistics Norway.

Keyfitz N (1985) *Applied Mathematical Demography*. Second edition. Springer: New York.

King R (2002) Towards a new map of European migration. *International Journal of Population Geography*, 8(2), 89–106.

Klabunde A and Willekens F (2016) Decision-making in agent-based models of migration: state of the art and challenges. *European Journal of Population*, 32(1), 73–97.

Knauth B (2011) Migration Statistics Mainstreaming. In: International Statistical Institute, Proceedings of the 58th World Statistical Congress, Dublin, 2416–2420.

Kőszegi B and Rabin M (2007) Reference-dependent risk attitudes. *American Economic Review*, *97*(4), 1047-1073

Koser K (1997) Information and repatriation: the case of Mozambican refugees in Malawi. *Journal of Refugee Studies*, *10*(1), 1-17

Lanzieri G (2019a) Towards a single population base in the EU. Meeting of the Group of Experts on Population and Housing Censuses, UNECE, Geneva, 18–20 September 2019.

Lanzieri G (2019b) An alternative view on the statistical definition of migration. Meeting of the Work Session on Migration Statistics, UNECE, Geneva, 29–31 October 2019.

Lucas RE Jr (1976) Econometric Policy Evaluation: A Critique. In: K Brunner and A Meltzer (eds.), *The Phillips Curve and Labor Markets*. Carnegie-Rochester Conference Series on Public Policy, 1: 19–46.

Mackie JL (1965) Causes and Conditions. *American Philosophical Quarterly*, 12, 245–265.

Makridakis S and Taleb N (2009). Living in a world of low levels of predictability. *International Journal of Forecasting*, 25(4), 840–844.

Massey DS, Arango J, Hugo G, Kouaouci A, Pellegrino A and Taylor JE (1993) Theories of International Migration: Review and Appraisal. *Population and Development Review*, 19(3), 431–466.

Massey DS, Arango J, Hugo G, Kouaouci A and Pellegrino A (1999) *Worlds in Motion: Understanding International Migration at the End of the Millennium*. Oxford: Clarendon Press.

Migali S, Natale F, Tintori G, Kalantaryan S, Grubanov-Boskovic S, Scipioni M, Farinosi F, Cattaneo C, Benandi B, Follador M, Bidoglio G, McMahon S and Barbas T (2018) International Migration Drivers. Ispra: JRC/ Luxembourg: Publications Office of the European Union.

Myrdal G (1957) *Rich Lands and Poor: The Road to World Prosperity*. New York: Harper and Bothers.

Neagu I and Schiff M (2009) Remittance stability, cyclicality and stabilizing impact in developing countries. Policy Research Working Paper 5077, Washington, DC: World Bank.

Nowok B and Willekens F (2011) A probabilistic framework for harmonization of migration statistics. *Population, Space and Place*, 17(5), 521–533.

O'Hagan A, Buck CE, Daneshkhah A, Eiser JR, Garthwaite PH, Jenkinson DJ, Oakley JE, and Rakow T (2006) *Uncertain Judgements – Eliciting Experts' Probabilities*. New York: John Wiley.

Pietzsch B, Fiedler S, Mertens KG, Richter M, Scherer C, Widyastuti K, Wimmler M-Ch, Zakharova L and Berger U (2020) Metamodels for Evaluating, Calibrating and Applying Agent-Based Models: A Review. *Journal of Artificial Societies and Social Simulation*, 23(2)9, http://jasss.soc.surrey.ac.uk/23/2/9.html

Popper KR (1982) *The Open Universe. An Argument for Indeterminism*. London: Hutchinson.

Ragin CC (1987) *The comparative method: Moving beyond qualitative and quantitative strategies*. Los Angeles: University of California Press.

Ramirez R, Mukherjee M, Vezzoli S and Kramer AM (2015) Scenarios as a scholarly methodology to produce "interesting research". *Futures*, *71*, 70-87.

Ravenstein EG (1885) The laws of migration. *Journal of the Royal Statistical Society*, 48(2), 167–227.

Raymer J, Wiśniowski A, Forster JJ, Smith PWF and Bijak J (2013) Integrated modeling of European migration. *Journal of the American Statistical Association*, 108(503), 801–819.

Rees P and Lomax N (2020) Ravenstein Revisited: The Analysis of Migration, Then and Now. *Comparative Population Studies*, 44, 351–412.

Rogers A (1975) *Introduction to Multiregional Mathematical Demography*. New York: Wiley.

Rogers A and Castro LJ (1981) Model Migration Schedules. IIASA Report RR8130. Laxenburg: International Institute for Applied Systems Analysis.

Sahota GS (1968) An Economic Analysis of Internal Migration in Brazil. *Journal of Political Economy,* 76 (2), 218–245. doi:10.1086/259395.

Saltelli A, Bammer G, Bruno I, et al. (2020) Five ways to ensure that models serve society: A manifesto. *Nature*, 582, 482–484.

Sardoschau S (2020) The Future of Migration to Germany: Assessing Methods in Migration Forecasting. Project Report #DPR 1|20. Berlin: DeZIM Institut.

Simon HA (1983) *Reason in Human Affairs*. Stanford, CA: Stanford University Pres

Shaw Ch (2007) Fifty Years of United Kingdom National Population Projections: How Accurate Have They Been? *Population Trends*, 128: 8–23.

Smith SK (1997) Further thoughts on simplicity and complexity in population projection models. *International Journal of Forecasting*, 13(4), 557–565.

Taleb NN (2007) *The Black Swan: The impact of the highly improbable*. New York: Random House.

Tetlock PE and Gardner D (2015) *Superforecasting: The Art and Science of Prediction*. London: Random House.

Tversky A and D Kahneman (1974) Judgment under Uncertainty: Heuristics and Biases, *Science*, 185(4157), pp. 1124–1131

United Nations (1998) Recommendation on statistics of international migration. Revision 1. Statistical Papers ST/ESA/STAT/SER.M/58/Rev.1. New York: United Nations.

Van Hear N, O Bakewell and Long K (2018) Push-pull plus: reconsidering the drivers of migration, *Journal of Ethnic and Migration Studies,* 44: 927-944.

Vishnevsky A and Zayonchkovskaya Zh (1994) Emigration from the former Soviet Union: the fourth wave. In: H Fassmann and R Münz (eds.), *European Migration in the Late Twentieth Century: Historical Patterns, Actual Trends and Social Implications*. Aldershot: Edward Elgar.

Wall M, Costello R and Lindsay S (2017) The miracle of the markets: Identifying key campaign events in the Scottish independence referendum using betting odds. *Electoral Studies*, 46(1), 39–47.

Willekens F (2018) Towards causal forecasting of international migration. *Vienna Yearbook of Population Research*, 16, 199–218.

Williams AM and Baláž V (2012) Migration, risk, and uncertainty: Theoretical perspectives. *Population, Space and Place*, *18*(2), 167-180.

Wilson T (2017) Can international migration forecasting be improved? The case of Australia. *Migration Letters*, 14(2), 285–299.

Wilson T and Shalley F (2019). Subnational population forecasts: Do users want to know about uncertainty? *Demographic Research*, *41*, 367-392.

World Bank (2019) *Migration and Remittances. Recent Developments and Outlook*. Washington DC: World Bank.

Zelinsky W (1971) The hypothesis of the mobility transition. *Geographical Review*, 61(2), 219–249.

■