WP 16.1
24 October 2013

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

STATISTICAL OFFICE OF THE EUROPEAN UNION (EUROSTAT)

Joint Eurostat/UNECE Work Session on Demographic Projections
organised in cooperation with Istat
(29-31 October 2013, Rome, Italy)

Item 16 –Population projections by age and sex and level of education

The scientific base of the new Wittgenstein Centre Global Human capital Projections: defining assumptions through an evaluation of expert views on future fertility, mortality and migration

Wolfgang Lutz, Wittgenstein Centre for Demography and Global Human Capital

Abstract

This paper proposes a new approach for defining the assumptions in population projections. It is based on a broad discussion and argumentation exercise which critically assesses the science basis of alternative arguments that relate to the forces which jointly shape the future trends in fertility, mortality and migration. It starts by discussing the results of a recent Eurostat survey among all EU national statistical offices in which there appears a clear consensus that the current practice of defining assumptions for population projections should be improved through stronger interactions with the scientific community. As an instrument for such interactions, the paper goes on to propose an interactive spreadsheet in which more than 100 arguments relating to the forces shaping future fertility, mortality and migration have been formulated. Invited experts then evaluate these arguments with respect to their validity and their potential impacts on the force under consideration. This interactive questionnaire has first been used to define the assumptions in the official UK population projections and later with an international group of 17 mortality experts. Most recently it has been applied for helping to define the fertility, mortality and migration assumptions for all countries in the world based on the input of over 500 international population experts.

1 Introduction

In times of accelerating social, economic and environmental changes, science-based information about the future becomes ever more important. The demand for such authoritative information often exceeds the supply of credible projections. There are many dimensions along which projections could be doing better than is the current practice. In the field of population projections, such improvements could be essentially along three lines:

(a) **Offering more detail** about the stratification of the population along dimensions other than only age and sex. The European Commission's 6th Framework Project "Bridging the Micro-Macro Gap in Population Forecasting" (MicMac) has pioneered approaches

which explicitly consider educational status, health status and household status in addition to the traditional dimensions of age and sex. These additional dimensions are assumed to cover important sources of heterogeneity for demographic behaviour which implies that their explicit consideration in projections will both improve the accuracy of the projections and provide more detailed information that is of direct use for planning and other purposes.

(b) **Providing more information about uncertainty**. While most users of projections are primarily interested in a best guess forecast, there is increasing demand for explicit consideration of the full range of uncertainty. Traditionally, this demand has been met by producing alternative "variants" or "scenarios" which are supposed to cover some "plausible" range of future trends. More recently, however, planning agencies explicitly demand probabilistic population projections in order to have a "demographic risk function" to be matched to their cost function. In terms of future pension entitlements, for instance, a very minor deviation from the base line projection of the proportion of the population above age 65 in 2030 implies additional government expenses or savings in the billions of Euros. This is also the reason why the UK Ministry of Finance recently explicitly requested a probabilistic population projection from the UK Office of National Statistics in order to enter it into its main economic model.

(c) **Provide a better substantive justification of the assumptions made**. The results of projections crucially depend on the specific assumptions made. As the survey of statistical agencies discussed below will describe, these assumptions are typically defined within those agencies after consultation with the experts. This is usually done in a rather informal way and the resulting choices tend to reflect "expert opinion" rather than transparent science-based reasoning which some of the users might expect. All of the statistical agencies state that this particular aspect of producing projections is the one most in need of further improvements.

This paper will primarily address the third issue concerned with the process of defining assumptions. But inevitably such discussions also touch upon the issues of uncertainty and that of population heterogeneity.

The paper begins with a summary of the results of a survey among all EU national statistical agencies about the current practice in population projections and their views of desirable further improvements. We will then step back and discuss some more theoretical meta-scientific points about the common fallacies associated with expert opinion and frequently stated arguments, and the possible science-based solutions. We continue by presenting a new argument-based questionnaire that was used in the most recent population projections of the UK and summarize the results of a more extensive exercise along these lines carried out in the context of a MicMac workshop on future mortality trends. The paper will conclude with a discussion of what we have learned and an outlook to the future of population projections.

2 Survey of current practices in national statistical agencies in the EU

Virtually all national statistical agencies in the world as well as inter-governmental agencies such as the United Nations and Eurostat have been producing regular population projections by age and sex following the so-called cohort-component projection method. The arithmetic of making cohort-component population projections has essentially remained unchanged since Edwin Cannan proposed it in 1895. However, the way in which demographic forecasters make their assumptions concerning future trends in fertility, mortality and migration is always potentially changing. This section reviews the current practice of statistical agencies in Europe (EU-25) for defining the fertility, mortality and migration assumptions in population projections, and draws preliminary conclusions from the review. Perhaps the most significant finding is that all national statistical offices that replied to a questionnaire on this topic agreed that there is a need for improvement in the methods used to make assumptions. In particular, the offices charged with making population projections would welcome more structured interactions with the demographic research community.

One of the first steps of the European Commission's MicMac project was to collect information on the current use of external experts in defining fertility, mortality and migration assumptions. The national statistical offices (NSOs) of the European Union countries were asked to provide information on what had been done during the production of the *most recent* population projections. Each office received a questionnaire from Eurostat – designed by the International Institute for Applied Systems Analysis (IIASA) – and 21 out of 25 national statistical offices returned the completed questionnaire by late November 2005. In addition to pre-set answers to the 12 questions, the questionnaire provided space for open-ended comments on each question. A complete account of these answers and comments has been provided by Prommer and Wilson (2006).

The aim of the questionnaire was to assess the current status of expert involvement and methodology in making population forecasts by the national offices. The second aim of the questionnaire was to evaluate what future improvements could be made in the process by which experts contribute to the definition of assumptions in population projections. The results of the survey are presented in Appendix Table 1 of Lutz (2009)[1].

The majority of the offices indicated that they generally use three future pathways for fertility, mortality and migration; some used only one or two. Only one office used stochastic methods to project future population. The most common approach is to create scenarios that cover a "plausible" range. The involvement of external experts and meetings are clearly important, but there is a marked gap between "old-15" who tend to have more elaborate ways of involving experts and "new-10" member states. If there are problems in finding a consensus on values, most offices make in-house decisions after consulting the experts. Three offices commissioned scientific studies from outside experts for the explicit purpose of helping with the definition of assumptions (though one of those does not publish its own forecasts).

Error analysis of past assumptions is also important. The decision to carry out either a systematic or a more qualitative analysis of past errors splits the respondents into two groups. However, no statistical office provided a description on the methodology they use for error

[1] Available here: http://www.iiasa.ac.at/publication/more_IR-09-037.php [September 26, 2013]

analysis. Half of the respondents define storylines (either combined for the three components of change or for each component separately) behind the assumptions. The other offices do not discuss storylines.

All national statistical offices agreed that there is a need for improvement in the methods used to make assumptions. Generally speaking, improvements in networking and in advancing the conventional methodology of scenario-based forecasts seem to have priority. The introduction of stochastic/probabilistic forecasting methods is not a high priority for most offices. It is not clear, however, if the slow pace of adoption of stochastic/probabilistic forecasts has to do with scientific criteria, or is simply due to the non-availability of human resources with the appropriate knowledge of the methodology. But there is evidence of lack of human recourses as three offices explicitly stated in question 12.

In the following we will have a closer look at the answers to some of the key questions. Anonymity of the specific quotes given is maintained throughout the text.

2.1 Number of alternative future paths defined for fertility, mortality, and migration

The majority of the statistical offices use the "scenario" design for the definition of the assumptions of the future paths of fertility, mortality, and migration. That means that the forecasters project the future population by defining at least one development path of future fertility, mortality and migration. This is commonly known as the "best-guess" or "most-likely" development. Roughly half of the central statistical offices prepare three different paths for each variable: fertility (52 percent), mortality (52 percent) and migration (43 percent). The second most common situation is that the NSOs use only one future path of each indicator; this is the case for five statistical offices. Two NSOs generated two different future developments for fertility, two NSOs for mortality, and four for migration. Generally speaking, the 3-3-3 version is most common, but there are some NSOs that use the 3-2-2, or the 3-1-1 composition. Only one statistical office identifies more than three possible future developments of mortality and migration, while two offices do so for fertility. Using stochastic population projections methods is still rare among the central statistical offices; only one institute stated that they use stochastic forecasting. Finally, one national statistical office replied that it did not officially publish projections and therefore uses the projections published by Eurostat.

The open-ended comments give us more insight how these assumptions are used in the projections. One statistical office that answered with a 1-1-1 assumption "produces demographic trend-calculation projection by using the so-called demographic component model, and does this for all municipalities". One office made national and regional projections for the "low-central and high hypothesis" in their latest projections. For another, the latest projections "were centred on the ventilation of those perspectives at the NUTS-3 level and calibrated them only on the former central hypothesis". Another office uses the 1-1-1 assumption hypothesis for short-term projections, but included two migration scenarios for long-term projections. Another NSO published the Eurostat baseline projection as the national variant, therefore, the answers in this questionnaire refer only to the procedures that took place during the discussion and harmonization of the assumptions for the three variants of population projections regarding fertility, mortality and migration for 2004 to 2050. One NSO that answered that they use the "standard" three variants (high, principal, low) also produce projections of a few special case scenarios, e.g., replacement fertility, no mortality improvement and zero migration.

2.2 Interpretation of the range defined by alternative assumptions

Thirteen national offices stated that the assumptions described cover a "plausible range". The office that uses the stochastic forecast methodology specifies intervals (usually 95 percent intervals) together with the type of distribution, etc. Two NSOs did not make any specification in the meaning of the alternative variants.

Some offices described the alternative variants, or how the "plausible range" was defined by filling in the open-ended comments: For instance, in one case the variants are considered as scenarios and extreme uncertainty and future international migration development is explained. Another NSO states that "… the high level of TFR (2.1) is explained by the fact that it is the symbolic value of the replacement level of fertility. As the baseline level was 1.8, the low level retained is 1.5 because it is 0.3 lower than the baseline (2.1 is 0.3 higher). It is as being the mean EU-15 level. For mortality, the various assumptions are explained by the expected trend of the future decrease in sex and age mortality rates (future trend is the same as past three decade trends / slow down in the trend at all ages / faster decrease at old ages). For migration balance, the level retained as baseline is explained as the mean level observed over the past two decades. A scenario 'without migration' is investigated but this assumption is only combined with the baseline assumptions on mortality and fertility and not with all assumptions (…). The high assumption on migration was 100,000 (+50,000 in comparison to the baseline), which was considered as a reasonable value by experts. The extra migration balance is supposed to consist in immigration only."

One office defined the standard variants as follows: "These are intended as plausible alternative scenarios and not to represent upper or lower limits for future demographic behaviour. For the special case scenarios we say: 'It is also sometimes useful to prepare special case scenarios, or 'what if' projections, to illustrate the consequences of a particular, but not necessarily realistic, set of assumptions.' "

There are variations of the definition when we consider the definition of another office. "For fertility the alternative assumptions are considered as covering a certain quantitatively specified uncertainty interval; for mortality alternative assumptions are considered as covering a 'plausible' range; and, for migration, behind the null scenario, the other assumption is based on values considered possible."

2.3 Who was involved in defining the assumptions

The baseline projection, or the "most-likely" or "best-guess" forecast, generally, is first discussed within the national statistical offices, and then discussed with outside experts (11 of 21 offices, or 52 percent of the respondents). The second most common approach is that the baseline forecast is initially proposed by experts and then discussed within the office (24 percent). Four NSOs (19 percent) discussed and defined it only within the office, and for one NSO the baseline projection was defined fully by outside experts. The one national statistical office that uses the Eurostat projections answered that both answers c) and d) applied. Despite appearances, this answer is not contradictory, as Eurostat produced the population projections in cooperation with the scientists of the country and the Central Statistical Bureau.

Typical comments on the procedure were: "The assumptions were drawn up by an expert group chaired by a member of the NSO and attended by outside experts as well as NSO experts." Or, the answer b) "is closest to the … situation. However, our initial meeting with six … academic experts was largely based on the assumptions used for the previous

projections and their views were part of the evidence we took into account in preparing new assumptions."

The absolute majority of the respondents (86 percent) answered that when they defined more projection variants for fertility, mortality and migration, they used the same mechanism to define the alternative scenarios. However, one bureau states that "the mechanism is not applied for each new forecast round for each component; this depends on whether new evidence is available or striking developments have taken place for a specific component."

2.4 Extent of involvement of external experts

This question refers to the number of experts involved in the decision process, in the number of consultations, and if separate meetings took place with different experts for fertility, mortality and migration. In total, 76 percent of the national offices answered that they involved external experts to define the future paths of the model determinants. Of those 16 institutes, 11 told us that they involved 10 or more outside experts; two institutes consulted between five and nine experts; and two offices one to four external experts. One bureau responded that "there is no fixed number of experts; mainly for migration".

The number of consultation meetings with such experts as a group is below 10 meetings in general. Some answers are vague, such as 1-10 meetings. Nine offices recorded fewer than five meetings. Eight of the 21 NSOs stated that there were separate meetings with different experts for fertility, mortality and migration, possibly explaining the relatively high number of meetings. Higher numbers of experts involved and a larger number of meetings were mostly seen in the statistical offices of the "old-15" countries of the European Union than in the "new-10" member states.

2.5 How did you deal with situations when experts disagreed

Three statistical offices told us that they were able to agree on common values, and hence they did not need any specific way to handle discrepancies. In cases where the experts could not reach a consensus for the model input, there is no clear policy. Around 50 percent of the national offices listen first to the experts but then take the final decisions in-house. Only three national statistical offices chose the way of majority vote, and two offices applied a variant of the "Delphi" method. To our understanding there is a clear need in the methods used to reach an agreement in case there are some problems among the external and/or internal experts. A vote is a fast procedure and is a good tool for generating quick, democratic decisions, but may not be applicable for scientific decision-making procedures.

One of the three bureaus that did not feel the need to deal with a "consensus mechanism" reacted as follows: "This was a common discussion to understand everyone's argument." Another NSO that answered with a) and d) answered that, "We base our assumption on facts and reasoning. We try to present as much [sic] facts as possible and we also try to specify where we are uncertain and show how and why we have decided in a certain way. We also try to describe the reasoning behind the assumptions thoroughly in the publication about the population projection." One office that answered that there was a majority vote, describes it in this way: "Assumptions are defined initially within the office. Papers are prepared to justify these assumptions. The experts usually agree with the basic assumptions." The in-house decision choice could be interpreted as the following possibility described here. "The discussion on assumptions should finish by selection [of] the most likely

variant for each component or by suggestion of verification. Conclusions from the meeting are implemented by NSO in calculation of official projection."

2.6 Definition of storylines, i.e., plausible visions of the future conditions that would result in certain trends

Over half the institutes (12 or 58 percent) do not use storylines to describe the reasoning behind the assumptions of the future paths of fertility, mortality and migration; seven institutes do use them. Two NSOs did not indicate any of the listed answers, and another office was unsure what was meant by "storylines". Five out of the eight offices that answered with yes define storylines for all three components; three of them defined consistent storylines to bind together at least two components. One of the three noted that they use both variants with more focus on the definition of the storylines for each component separately. The second didn't specify, and the third defined the storylines for low fertility and mortality jointly.

One NSO defines in detail the "binding together" of the storyline and thus provides insights into what probably happens in many of the NSOs. "The main variant is based on what are considered to be the most probable, and therefore reliable, trends for the future: a further fall in mortality, a slight increase in period fertility, inter-regional migration showing a constant probability, international migration at around levels experienced in the '90s. In addition to the main variant, two alternative scenarios have been considered regarding the development for each demographic component. The two alternative assumptions are intended to define the range of variation within which the future population will develop. The scenario imagined in the low variant is marked by minimal economic growth and limited attention paid to social problems. Given such a context, improvements in life expectancy would slow down and there would be no recovery in the fertility rate. Regarding migration, inter-regional and international flows would exhibit modest levels, resulting in a kind of stagnation owing to the low level of attraction exerted by the destinations concerned. Such a scenario would give rise to the lowest projected population level, characterised by the most unbalanced age structure. In the high variant the scenario assumes lively economic growth, providing the opportunity to increase investments also in the social and health fields. This would lead to a higher life expectancy than in the main variant and a considerable recovery in fertility rate. Furthermore, this scenario is also marked by a more intense population movement among regions and an increase in the attractiveness of … as a destination for immigrants from abroad. All of the foregoing factors would lead to the highest projected population level together with a more balanced age structure."

One office describes the need to define separate storylines for all three variables: "On some aspects the storylines are consistent, example: Many young [immigrants] … in the country look for a partner in their country of origin. The rising number of young … will have an upwards effect on the number of marriage migrants. As marriage migrants are rather traditional in behaviour these migrants will have an upwards effect on the fertility level of the … in the country, and of course also on overall fertility."

2.7 Expressed need for further improvements in the procedure for defining the assumptions and directions for improvement

There is a clear common agreement among all respondents: There is a need for improvement to define the assumptions on future fertility, mortality and migration. All 21 NSOs answered that they think further improvements in the procedure are required.

Only one office appeared to be somewhat hesitant about this choice in the questionnaire, but then stated: "We are satisfied with our procedures, but of course improvements can always be made, so we are continuously looking for improvements." Suggested improvements are: "In the sense that the analysis done on the data and the justification of the options should be more released in the publication." And "Everything can be improved. We are always open to discuss improvements. However we investigate new procedures carefully before we accept them as an improvement."

In what direction do the NSOs think the improvements should go? Here the tendency is fairly clear: The most widely chosen improvement options were to "have a more systematic review of all the substantive arguments behind the assumptions" (13 NSOs) and "have some structured interactions with the European demographic research community about the state of the art in our knowledge about future demographic trends (13 NSOs). This was followed by the desire to "involve more experts" (11 NSOs). Further down in the ranking came the hope for more exchange with other NSOs (9 NSOs) and with Eurostat (8 NSOs).

In a nutshell, this extensive and very informative enquiry among EU national statistical offices makes it very clear that all national offices think that the current practice for defining the assumptions for population projections is sub-optimal and needs further improvement. The results clearly indicate the direction into which to move on from the current practice: **Have a more systematic review of the substantive arguments behind the assumptions in the form of a structured interaction with the demographic research community which also facilitates the involvement of more experts**.

In the rest of this paper we will propose a new procedure for defining the assumptions that would try to achieve exactly this goal in all its dimensions. Furthermore, it would not only address the three top ranked directions of improvement but also, if conducted as a Europe-wide exercise in the future, have the potential to significantly strengthen the substantive collaboration among NSOs and with Eurostat.

3 Problems and fallacies associated with expert opinion and the road toward argument-based forecasting

Given the prominence and importance of the assessment of likely future demographic trends in Europe, it is surprising to see how little systematic attention the scientific community has been given to the evaluation of arguments underlying the assumptions of future fertility, mortality and migration trends. While the above-described survey shows that the National Statistical Offices put great hope in input from the demographic research community, this topic has largely been confined to the described processes within statistical agencies. Since these offices are in a way forced to make choices on assumptions in order to fulfill their mandate of producing population projections, they cannot escape this challenging task as easily as academics seem to be able to.

The users of population projections – which are by far the most visible and most relevant products that the demographic research community provides to the rest of society – would rightly expect a broadly-based discussion of likely future trends to be the main topic of international population meetings. They would also expect governments and intergovernmental bodies, which greatly rely on the accuracy of population projections in their policy formulations, to commission major studies to make sure that they get the best possible information about likely future demographic trends. But in stark contrast, the reality shows that at scientific meetings in the field of demography as well as in government-

sponsored activities around Europe, the discussion of assumptions used in projections is largely absent or at best a marginal topic.

It is important to point out here that the absence of such structured and prominent discussions is not due to the fact that we know all about the future. Quite the contrary, the sense of uncertainty about whether fertility in Europe will recover or continue to decline, or whether we are already close to a maximum life expectancy or will see continued increases, seems to be even higher than in the past. Moreover, studies on the accuracy of past population projections produced since the 1960s have shown that significant errors were made particularly with respect to anticipating the speed of population ageing in Europe. Generally, for most European countries, the national statistical agencies as well as the United Nations Population Division have assumed far too high fertility levels and far too low gains in life expectancy. While these two independent errors tend to cancel out when one is only interested in population size (fewer than expected deaths compensate the effect of fewer than expected births), they strongly reinforce each other when it comes to population ageing (ageing is enhanced by higher life expectancy and by lower fertility). One could even argue that these significant errors of past projections which failed to anticipate the actual speed of ageing have contributed to the fact that today's societies are not as well prepared for ageing (e.g., in terms of pension systems) as one could have hoped.

Given this situation, the approach presented in this paper will try to show the way for a better inclusion of available scientific knowledge into the process of defining the range of assumptions on future fertility, mortality and migration levels. In other words, it will attempt to facilitate the translation of the vast body of relevant research that exists in the demographic community as well as other related research communities into a definition of specific sets of science-based assumptions for projections. This follows up on earlier work by Lutz, Saariluoma, Sanderson and Scherbov (Lutz et al. 2000), on which the following section partly draws.

As shown in the previous section, up to now this process has mostly happened through the collection of expert opinions. Such procedures typically follow the tradition of Delphi methods that have been well developed and extensively documented elsewhere (Linstone and Turoff 2002). But the problems with expert opinions is precisely that they tend to be opinionated; this can result in all sorts of biases and distortions that are not desirable and do not necessarily reflect the best state of the art in the field. There is abundant evidence that experts tend to hold strong beliefs about the future, which are at the level of emotions and intuitions. Hence, the approach proposed in this paper is nothing short of trying to go beyond opinion-based Delphi and suggest a more objective science-based way. Of course, whenever one has to rely on the views of people in one way or another, this cannot be fully objective, but one can move into this direction by making it inter-subjective and applying the standard scientific tools of peer review and critical evaluation. But in order to make progress in this direction there needs to be something on the table to be evaluated and analytically reviewed. Hence, the argument-based approach will put specific arguments on the table that are directly relevant for the future course of the demographic force under consideration and which can then be critically assessed.

The following considerations are the product of an interdisciplinary collaboration between demographers involved in population projections and an experimental psychologist working in the field of cognitive science, in a way an expert on experts.

One important contribution that meta-science can make to any scientific approach is to investigate the problems in the way arguments are built in specific scientific fields. Such work can help the applied scientists find a more analytical way of thinking in their own fields. One

may ask why is it important to critically inspect the argumentative basis of a science. The logic of the answer is very straightforward: All scientific argumentation ends somewhere and from that point on, the area of intuitive assumptions begins. Infinite chains of arguments are impossible, but we need to be aware of this and reflect on the point when we choose to end the chain of argumentation. This point can be right next to the object of observation, in which case there is no argumentative foundation at all. It can also be too far away from the object in which case the arguments considered and the objects are hardly linked any more. The choice of this cut-off point needs to be based on expert judgment. But this is judgment at the meta-level rather than at the level of the object itself. Such judgment must be based on some sense of plausibility or intuition as it is typically called in cognitive science and foundational analysis.

Intuitions in the foundations of scientific ways of thinking are unavoidable. We cannot get around them; we have to learn how to live with them. The first step in this direction is to understand them in the right manner. The problem with intuitive foundations of science is not that all our intuitions would immediately and necessarily be false, but that we do not know whether they are true and to which degree they are true. This means that we have to adopt a dynamic stance toward them. We have to turn our attention to them and carefully consider the possible strengths and weaknesses in them. When we understand the intuitive foundations better, we are able to use this new understanding for the advancement of science. We can open new perspectives to knowledge and justify the search for new types of knowledge. Indeed, the ultimate goal of such foundational work is to deepen our understanding of what we are doing. This is a way to speed up the progress in science in general and in the field of making necessary assumptions for population projections in particular.

In the following, we will critically review some of the most common problems with expert judgment and the reference to empirical findings and discuss how arguments should be framed in order to avoid such problems. The goal is not to prove such arguments incorrect or empty in content, but rather to make sure that the specified arguments actually refer to possible causal mechanisms and are specific enough to be falsifiable. Only the evaluation of such arguments will add to our science-based knowledge about likely ranges of future demographic trends.

An important prerequisite for valid argumentation is the clarification of what is the assumed cause and what is the effect. The *explanans*, i.e., the explanatory premise, refers to statements that explain the *explanandum*, i.e., the phenomenon which should be explained on the grounds of *explanans*. In argumentation analysis, it is always central to consider carefully the form and explanatory power of the *explanans*.

A typical example for the confusion between *explanans* and *explanandum* is the assumption that something will not happen because it has not yet been observed, something the literature calls “curve illusion.” In this kind of false argument, one views the shape of an observed curve (the phenomenon to be explained) as the driver that produces a pattern. An example from the field of fertility assumptions is the frequently held view that there is some “rock bottom” fertility level below which fertility will not fall. This is simply justified by the fact that fertility has never fallen below such a level in any country. There may well be good arguments to assume that fertility will not approach zero in the future, but they cannot be based merely on the description of the “curve” observed so far. Interestingly, in the field of projecting life expectancy, although human history has never experienced a national life expectancy of above the current maximum of some 86 years for Japanese women, few people think that future increases will be impossible. But while a historically unprecedented level is no longer used as an argument in making mortality assumptions, the current practice of (blind) trend extrapolation is not much better from a meta-scientific perspective, if it does not

provide any plausible reasons for why life expectancy is assumed to continue to grow at the same speed as in the past or at a decelerating speed, as some agencies assume. More generally, in order to avoid such circularity of taking the *explanandum* for the *explanans*, one would have to anchor the argument in the world outside of the curve itself (the observed trends). If such an anchoring is not explicated, the argument cannot be valid.

The circulatory problem also exists when we refer to two different measurements which may be affected by the same cause but do not influence each other. When we measure a fever in a child, we do not think that the high temperature in itself is the illness or that the temperature in the mouth is caused by the temperature in the armpit, even though the correlation would be substantial. Instead, we look for the illness in the body, which explains the high fever measured at both points. We know that the body defends itself from many different types of illnesses by producing a fever and therefore, we look for further symptoms to cancel out incorrect diagnoses and to find the true explanation.

Another problem in this context of defining valid arguments is the confusion of differentials with causes. Much of the social sciences have been inspired by the observation of differentials. Individuals and their behavior differ from place to place, over time and among individuals. These differentials typically give rise to the formulation of explanations as to why the observed patterns of behavior differ. These explanations point the way to the more general causes of behavior. In many cases, however, the analysts stop short of providing real explanations for the observed differentials and suffice by describing only the differentials. An example for such inference from differentials to causes is when people point at the fact that urban women typically have lower fertility than rural women and conclude from this that increasing urbanization will lead to lower fertility. But this conclusion is only correct if it is assured that there is indeed a causal relationship from the kind of living environment on the number of children. There probably are such real causes, but in order to make it a valid argument, the possible causal mechanism has to be identified and discussed. This does not necessarily mean that they have to be proven in the sense of strong causality which may be very difficult. But at least the identified mechanisms should result in a plausible storyline. Such an argument can then be properly evaluated both with respect to its validity and its relevance.

The same problem affects the currently popular notion of a “second demographic transition” (SDT). It is a name given to a bundle of observed trends in certain values related to sex and partnership and is by its very nature an *explanandum* (or a “curve” in the above terminology). Although SDT is sometimes referred to as a theory, it does not potentially have predictive power (such as stating that a country that will move into the direction of a more liberal attitude toward sex will have fewer children in the future) and hence must not be mistaken as a testable *explanans*. In this respect the concept of SDT also suffers from another frequent problem that makes many proposed patterns of explanation inappropriate as valid arguments, namely, the lack of specificity or, in other words, the fact that they are too general. If an argument is too vaguely formulated or too broad and general in its content so that there is no way to potentially reject it, the argument is not helpful for broadening our science-based understanding of the future.

On the opposite end of the spectrum of problems lie those with arguments that are too specific and too narrow. While such arguments may well be falsifiable in the sense that they have specific information content that can be evaluated in the light of empirical evidence and theoretical cohesion, they may not add much to our overall understanding of likely future trends because they only address a very narrow aspect of all factors that jointly determine the future trend of the demographic force under consideration. A good example for such an argument in the field of fertility determinants is a focus on declining human sperm counts.

While there seems to be convincing evidence that in some countries there have been significant declines in the quantity and quality of sperm counts, and a sufficient number of healthy sperm clearly are a prerequisite for natural conception, some commentators have taken this as an explanation for the declining birth rate. But this relationship is far from straightforward. As discussed in a recent special issue of the *International Journal of Andrology* (Jørgensen et al. 2006), declining sperm quality and counts may well affect the waiting time to conception (and only in rare cases lead to infertility), but this interacts in a complex way with characteristics of the partner as well as the nature of the partnership.

But the problem of partial explanations is much broader than the example above. Essentially all arguments about future trends in fertility, mortality and migration focus on certain partial aspects, while leaving others out. Hence, it is one of the most challenging tasks for the development of a new model for argument-based assumption making to bring these different aspects together in a comprehensive way in which the relative importance of the different arguments in determining the future course of the force under consideration are assessed. This will be done in the form of weights to be attached to the different factors that should resemble reality as closely as possible. In other words, we will distinguish between assessing the validity of certain arguments and their relevance in terms of influencing the overall trend of the demographic force.

Based on the above-described considerations and the identification of possible traps and pitfalls in the specification of arguments, in the following section we will present a scheme of core substantive arguments that try to avoid (as far as possible) the above problems. These could become the basis of a systematic future scheme for defining argument-based assumptions for population projections for essentially all countries in the world. While the specific formulation of forces and arguments in the following section is geared toward population projections in industrialized countries, it will be relatively easy to adapt the framework for use in developing countries as well.

4 Recent application to a new set of global population projections produced by the Wittgenstein Centre for Demography and Global Human Capital

"Can knowledge improve forecasts?" was the title of a famous paper by Nathan Keyfitz (1981), in which he expresses the view that demographic trends are easier to forecast than many social and economic trends which are often seen as drivers of fertility and mortality. But how should the demographic trends themselves be forecast? What should be the basis for assumptions about future fertility, mortality, and migration trends? Ahlburg and Lutz (1999), in the introduction to a special issue of PDR on the topic "Frontiers of Population Forecasting" (Lutz et al., 1999), interpret this view of Keyfitz (after discussions with Nathan Keyfitz at IIASA) by suggesting that demographic trends should not be entirely derived from other forecasts (as in the World 3 Model of the "Limits to Growth" by Meadows et al., 1972), or based on blind extrapolation or replication of past trends. Instead, they suggest summarizing the scientific community's knowledge base concerning future demographic trends through a structured process of expert solicitation. This view was also inspired by the influential work of Armstrong and colleagues on forecasting outside the realm of demography (Collopy and Armstrong, 1992), which demonstrated that structured judgment outperforms either judgment alone or a statistical model alone (Ahlburg and Lutz, 1999). This insight guides the approach chosen in this study.

In parallel to the scientific discussion, the practice in statistical offices has been moving in this direction of using structured expert judgment for defining the assumptions. Virtually all national statistical agencies in the world, as well as inter-governmental agencies such as the United Nations (until recently) and Eurostat, have been producing regular population projections by age and sex, following the cohort-component projection method with assumptions on future fertility, mortality and migration based on expert judgement

It is worth noting here that the United Nations Population Division (UNPD) recently (since the 2010 assessment) decided to go into a very different direction which bases assumptions largely on a statistical model using only past national level time series within the context of a particular structure and disregards whatever substantive knowledge there exists in the international scientific community about the country- and region-specific factors influencing future fertility, mortality and migration trends. A detailed comparison of the approach chosen in this volume with the new UN approach is currently in progress and will be published elsewhere. At this point is suffices to say that the two approaches follow quite different forecasting philosophies. While the new UN approach essentially assumes that the best we can assume is that the future will see a replication of past trends interpreted within the framework of their model – which essentially makes the entire demographic research community redundant when it comes to producing projections – we chose to follow the alternative path of making the substantive arguments upon which our forecasts rest explicit. We did this by improving the structured procedures for soliciting country-specific expert knowledge and significantly enlarging the number of experts involved in the process to over 500. And as will be describe in detail in the following chapters the substantive expert assessment of alternative arguments relevant for the future trends will be blended with formal statistical models much in the spirit recommended by Collopy and Armstrong (1992).

In this section we first review the current practice of statistical agencies in the European region and note a survey that indicated near consensus on the need for improving the procedures by which expert knowledge is assessed as the basis for making assumptions. The main results of the survey was that the offices charged with making population projections would welcome more structured interactions with the demographic research community. We then discuss the proposal for a systematic argument-based approach to making demographic assumptions that was developed under the European Commission's "MicMac" project ("Bridging the micro-macro gap in population forecasting", www.nidi.nl/micmac) and has been operationalized for the projections presented in this volume.

In 2005, all national statistical offices (NSOs) of the European Union countries were asked to provide information on their procedures for producing their most recent population projections. Each office received from Eurostat a IIASA-designed questionnaire, which 21 out of 25 national statistical offices returned. The results of this survey were documented in Prommer and Wilson (2006) and Lutz (2009). The most common procedure was to create scenarios that cover a "plausible" range of future fertility, mortality, and migration. The involvement of external experts and meetings was generally considered very important, but all national statistical offices suggested that the methods used to make assumptions could use further improvements.

As the most desirable procedure, one NSO explicitly stated: "We base our assumption on facts and reasoning. We try to present as many facts as possible and we also try to specify where we are uncertain and show how and why we have decided in a certain way. We also try to describe the reasoning behind the assumptions thoroughly in the publication about the population projection."

The NSOs were also asked about the best directions for such improvements. The following three improvement options were most frequently endorsed: (a) “have a more systematic review of all the substantive arguments behind the assumptions” (b) “have some structured interactions with the European demographic research community about the state of the art in our knowledge about future demographic trends”, and (c) “involve more experts”.

In a nutshell, this extensive and informative enquiry among EU national statistical offices shows that all national offices consider the current practice for defining the assumptions for population projections as sub-optimal and needing further improvement. The results further indicate the direction in which to move: Have a more systematic review of the substantive arguments behind the assumptions in the form of a structured interaction with the demographic research community, which also facilitates the involvement of more experts.

This important view from the practitioners of population projections, together with the theoretical insights discussed above provided the basis for the new expert-argument based approach that was developed at IIASA and has now been broadly applied to defining the assumptions for the set of global population projections presented in this volume. This process involved more than 500 population experts from around the world who answered the online questionnaire or participated in one of the five Meta-expert meetings. Before describing the design and conduct of this exercise, it useful to step back and discuss some of the general problems associated with any kind of expert solicitation regarding future trends, and with the general reluctance of academics to make statements about the future.

The meta-science of expert knowledge

Given the public prominence and substantive importance of likely future demographic trends in virtually all countries of the world, we find it surprising how little systematic attention the scientific community has given to the evaluation of arguments underlying the assumptions of future fertility, mortality and migration trends. While the above-described survey shows that national statistical offices place hope in input from the demographic research community, this topic has largely been confined to the statistical agencies. Since these offices cannot avoid making choices about assumptions in order to fulfill their mandate of producing population projections, they cannot escape this challenging task as easily as academics seem to be able to. Our approach attempts to translate the vast body of relevant research in the demographic community and other related research communities into specific sets of science-based assumptions for projections. This follows earlier work by Lutz, Saariluoma, Sanderson and Scherbov (2000), that explicitly discusses some of the shortcomings of expert based procedures. Such procedures typically follow the tradition of Delphi methods that have been developed and extensively documented elsewhere (Linstone and Turoff, 2002).

The main problem with expert opinions is that they tend to be opinionated; this can result in undesirable biases and distortions that do not necessarily reflect the state of knowledge. There is abundant evidence that experts tend to hold strong beliefs about the future that are at the level of emotions and intuitions (Lutz, 2009). Hence, the approach proposed here goes beyond opinion-based Delphi to embody a more objective science-based procedure. Of course, whenever one relies on the views of people, the result cannot be fully objective, but one can move toward objectivity by making the process inter-subjective and applying the standard scientific tools of peer review and critical evaluation. In order to make progress there must be something on the table to be evaluated and analytically reviewed. Hence, our argument-based approach puts specific arguments on the table that are directly relevant for the future course of the demographic force under consideration and which can be critically assessed.

In designing this new argument-based approach, we collaborated with an experimental psychologist working in the field of cognitive science, who is an expert on experts. We learned that one important contribution that meta-science can make to any scientific approach is to investigate problems in the way arguments are built in particular scientific fields. Such an investigation can help applied scientists find a more analytical way of thinking in their own fields. It is important to understand that all scientific argumentation ends somewhere, and from that point on the operation of intuitive assumptions begins. Infinite chains of arguments are impossible. Being aware of this, one should reflect on when to end the chain of argumentation. This point might be right next to the object of observation, in which case there is no argumentative foundation at all. It can also be too far away from the object, in which case the arguments considered and the objects are linked only tentatively. The choice of this cut-off point needs to be based on expert judgment. But this is judgment at the meta-level rather than at the level of the object itself. Such judgment must be based on some sense of plausibility or intuition, as it is typically called in cognitive science and foundational analysis. Hence, intuitions in the foundations of scientific ways of thinking are unavoidable. We cannot get around them; we have to learn to live with them, and the best way to do so is to be cognizant of the known traps and fallacies and to be entirely transparent about the choices made.

The systematic inquiry of expert arguments chosen for this study

Based on these considerations and principles, a team at IIASA and the Vienna Institute of Demography (VID) designed an on-line questionnaire with separate segments for fertility, mortality, and migration. The questionnaire was developed gradually, first only with respect to mortality when it was tested with a group of 17 mortality experts in the context of the above mentioned MicMac project (Lutz, 2009). It was then further operationalized and programmed in Excel by the National Office of Statistics in the United Kingdom, which used it for systematically collecting expert-based assessments on fertility, mortality, and migration for t population projections. A team at IIASA further developed the questionnaire for its broader use for all countries. For each of the three segments the arguments were grouped according to selected major forces that were defined to be as independent from each other as possible.

For fertility and mortality six such major forces were identified, and for international migration five. They are:

Major forces on which future fertility (F) will depend:

F1. Changing cultural and social forces in fertility ideals, norms and desires

F2. Changing patterns of partnerships and gender differences

F3. Changing roles of government policies (child care facilities, housing, etc.)

F4. The changing environment in terms of employment and the economy

F5. Changing bio-medical conditions (sperm quality and counts, female fecundability, new methods for assisted conception, etc.)

F6. Changes in educational attainment

Major forces on which the future of life expectancy (L) will depend:

L1. Progress in biomedical technology

L2. Changes in the effectiveness of health care systems

L3. Changes in health-related behavior

L4. Possible new infectious diseases and resurgence of old diseases

L5. Environmental change, disasters and wars

L6. Changes in population composition and differential trends in population subgroups

Major forces influencing international migration (M):

M1. Trends in economic development as a driver of international migration

M2. Climate change and conflict as possible drivers of international migration

M3. Demographic factors as a possible driver of international migration

M4. Changes in the costs of migration (in the broader sense)

M5. Trends in migration regimes and policies

For each of these 17 major forces up to 10 specific arguments were listed, with responding experts invited to add additional self-defined arguments.[2] For each of the arguments the experts were asked to judge the degree of correctness based on the scientific evidence. They were given five predefined choices about the validity of the argument ("very likely to be right", "more right than wrong", "do not know/ambivalent", "more wrong than right", "very likely to be wrong"). As can be seen from Figure 1, which provides a screen shot of one of the argument assessment pages, these answers were translated into a numerical factor (in parentheses below) ranging from 0.0 for "very likely to be wrong to" to 1.0 for "very likely to be right".

Whether the argument is right is not all that matters; the relevance of the argument for influencing the future course of the force under consideration is also important. Some arguments may likely be true, but completely irrelevant for the question under consideration. Hence, experts were asked a second question concerning the likely impact of each argument on the demographic component (see Figure 1), which says, "Regardless of your answer above, if the above argument were completely true, what effect would this have on the future of …". Again the user could choose among five answers ranging from "strongly decreasing" (assigned a factor of -1.0) to "strongly increasing" (assigned a factor of +1.0). If the respondent believes that the factor has no effect on the demographic outcome considered, then this answer is assigned to zero.

In a final assessment, the two factors stated for any given argument are multiplied, with the total argument impact score automatically shown, as in the scale at the bottom of Figure 1. This multiplicative assessment of the degree of the statement's correctness and of its impact if correct can thus range from strong negative impact, given by -1.0 (when the statement is considered correct and to have a strong negative impact), to a strong positive impact given by +1.0 (when the statement is considered incorrect and to have a strong positive impact). If the statement is either considered to be wrong or right but having no impact the total impact score (resulting from the multiplication of the two scores) is zero.

In mid-2011 all members of international population associations (IUSSP, PAA, EAPS, APA and all other regional population associations) were invited to participate in the online survey. More than 550 responses were submitted. The lead authors of this volume, most of whom are affiliated with the Wittgenstein Centre, systematically analysed the responses and led a series

[2] The full questionnaire can be found at http://webarchive.iiasa.ac.at/Research/POP/ExpertSurveySandbox/

of five Meta-Expert workshops to review and focus them. These specialized workshops consisted of groups of 8-12 leading experts in the respective fields. Between October 2011 and February 2012 the meetings were held on five continents: Migration (Boulder, Colorado.), Low Fertility (Vienna), High Fertility (Kathmandu), Low Mortality (San Jose, Costa Rica) and High Mortality (Cape Town). Based on these meetings, we defined alternative education-specific fertility, mortality and migration assumptions for all countries, as is described in detail in other papers in this session.

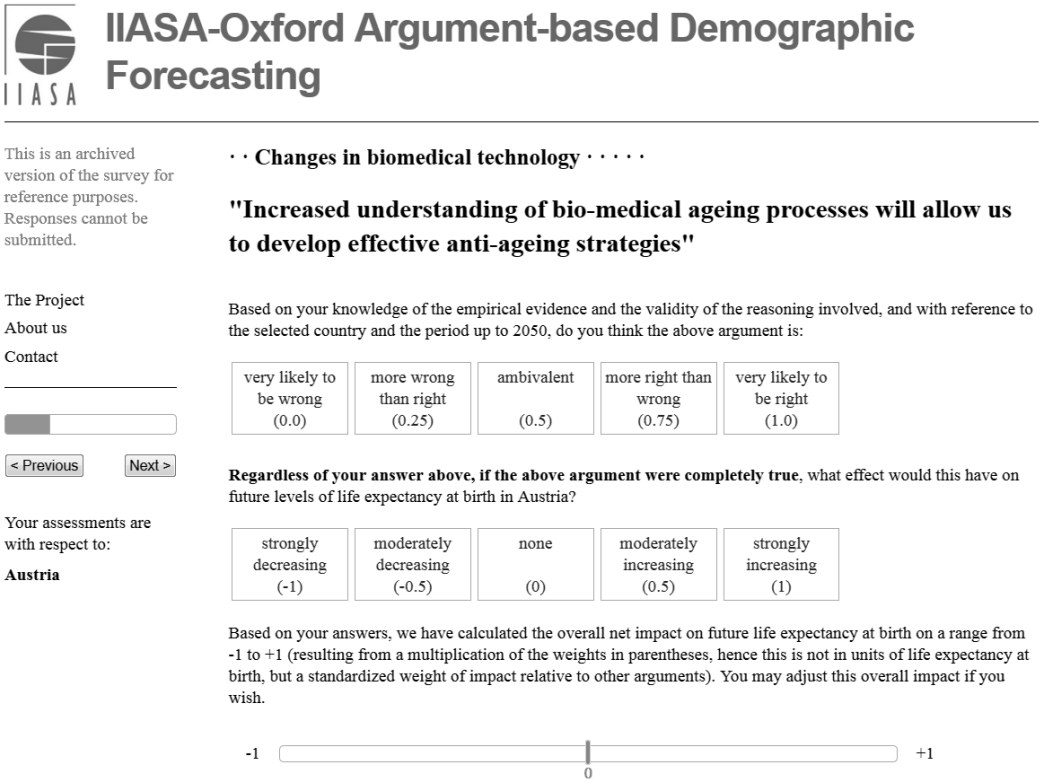

Figure 1: screen shot from a sample page of the online questionnaire listing one argument on future life expectancy under the force “Changes in biomedical technology”.

5 Discussion and Outlook

This new argument-based approach for providing a more scientific basis for defining assumptions for population projections is directly responsive to the needs and priorities as expressed in the above-described survey of EU National Statistical Offices. It attempts to facilitate the desired, more systematic review of the substantive arguments behind the assumptions in the form of a structured interaction with the demographic research community, which also allows for the involvement of many more experts. It offers an option for the future for all national statistical agencies in Europe and, if conducted in a concerted effort, it can facilitate the interaction among different agencies and with Eurostat. The UK Office of National Statistics has already used this questionnaire for two rounds of projections and further improved it. Recently, a further modified version of the questionnaire was sued for a major exercise to produce new population projections for all countries in the world carried out by the Wittgenstein Centre for Demography and Global Human Capital, a collaboration

between IIASA, the Vienna Institute of Demography of the Austrian Academy of Science and the WU (Vienna University of Economics and Business).

The application of this argument-based forecasting approach is not limited to applications for fertility, mortality and migration. An isomorphic procedure can easily be applied to other demographic dimensions considered in the MicMac model such as education, health status or household status. It is an appropriate way to go beyond rather simple ad hoc scenarios in the definition of future demographic transition rates and exploit the entire body of existing literature and knowledge about the substantive forces that are likely to determine these transitions in the future. One may choose to further decompose the trends under consideration. In the case of mortality, for instance, one may consider it desirable to separately study the trends in different causes of death. This would be feasible in the context of the described approach although one would have to be careful not to increase the complexity beyond a manageable point.

Another potentially consequential innovation of this approach is that the standardized computer-based format of the questionnaire allows for the inclusion of almost unlimited numbers of experts. The questionnaire can be widely distributed electronically and certain quality criteria can help to make sure that only the information provided by competent persons is taken into account. For this the first criterion would clearly be the ranking of competence on the subject matter and the statement about the years already working in the field as provided by the respondent. It would also be advisable not to invite every chosen expert to answer all three sections on fertility, mortality and migration, but to choose only the field in which he/she feels most expert. It would be easy to "test" the competence of the respondent in the chosen field by including a few factual questions or questions relating to the key literature in the field to filter out those who do not know the field well enough. Since the processing of the responses will be anonymous, the "failing" respondents would not know about this and hence there would be no hurt feelings. As already indicated above, certain criteria applied to the internal consistency of the answers given could be used to assign different weights to different experts when aggregating the answers across experts.

The inclusion of a larger number of experts who are not only drawn from academic demography, but also from a broader range of other relevant disciplines as well as from other government agencies and civil society organizations, might serve two additional important goals: (1) it could bring up some more unconventional but relevant information and views that the traditional experts do not see in a possibly too narrow perspective, and (2) it could create a broader public interest and therefore ownership of population projections in society. Once population projections are not seen as something cooked up behind closed doors by a few people, but rather convey the sense that a large number of persons from different segments of society have contributed to them, more people will feel ownership of them and possibly take the projections more serious in their own work. And last but not least, more people will get a better sense of what is known and what is unknown about the future demographic trends once they have to work through all the argumentation process themselves. It may also sensitize them as to what factors can be changed by policies and what has to be taken as given.

This points to another possible use of the results of such an argumentation exercise that goes beyond forecasting and has policy relevance In this sense the argument-based approach presented here not only offers a better scientific basis for projections than current practice, it also offers otherwise unavailable insights into the relative importance of the different factors that influence the different forces which jointly shape our demographic future.

6 References (not yet complete)

Jørgensen, N., C. Asklund, E. Carlsen, and N.E. Skakkebæk. 2006. Coordinated European investigations of semen quality: Results from studies of Scandinavian young men is a matter of concern. *International Journal of Andrology* 29: 54-61.

Linstone, H.A. and M. Turoff, Eds. 2002. *The Delphi Method: Techniques and Applications*. Newark, NJ: New Jersey Institute of Technology. Digital reproduction of the original 1975 version: http://is.njit.edu/pubs/delphibook/

Lutz, W., Ed. 1991. *Future Demographic Trends in Europe and North America. What Can We Assume Today?* London: Academic Press.

Lutz, W., Ed. 1994. *The Future Population of the World. What Can We Assume Today?* London: Earthscan.

Lutz, W., Ed. 1996. *The Future Population of the World. What Can We Assume Today?* Revised edition. London: Earthscan.

Lutz, W. 2009. *Toward a Systematic, Argument-Based Approach to Defining Assumptions for Population Projections*. Interim Report IR-09-037. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Lutz, W., P. Saariluoma, W.C. Sanderson, and S. Scherbov. 2000. New Developments in the Methodology of Expert- and Argument-Based Probabilistic Population Forecasting. Interim Report IR-00-020. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Lutz, W., W. Sanderson, and S. Scherbov. 1997. Doubling of world population unlikely. *Nature* 387: 803-805.

Lutz, W., W. Sanderson, and S. Scherbov. 2001. The end of world population growth. *Nature* 412: 543–545.

Lutz, W., W.C. Sanderson, and S. Scherbov, Eds. 2004. *The End of World Population Growth in the 21st Century: New Challenges for Human Capital Formation and Sustainable Development*. London: Earthscan.

Lutz, W., W. Sanderson, and S. Scherbov. 2008. The coming acceleration of global population ageing. *Nature* 451: 716-719.

Lutz, W., Butz, W., and S. KC. 2014 forthcoming. World population and human capital in the 21st century. Oxford University Press.

Nakićenović, N. et al. 2000. *Emissions Scenarios*. A Special Report of Working Group III of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press.

Prommer, I. and C. Wilson. 2006. An Analytical Summary of the Current Practices of Definition of Assumption Making in Population Projections. Deliverable D14. Report prepared for the European Commission 6th Framework Project "Bridging the Micro-Macro Gap in Population Forecasting" (MicMac), Contract No. SP23-CT-2005-006637. Laxenburg, Austria: International Institute for Applied Systems Analysis.