

Distr.: General  
07 November 2019

English only

---

## **Economic Commission for Europe**

Conference of European Statisticians

### **Work Session on Demographic Projections**

Belgrade, 25–27 November 2019

Item No. 4 of the provisional agenda

#### **Methodology**

## **Using expert elicitation to build long-term projection assumptions**

**Note by Statistics Canada\***

### *Summary*

Expert judgment is valuable when there is either a lack of good data, insufficient knowledge about underlying causal mechanisms, or apparent randomness in trends. Most statistical agencies consult with experts in some manner prior to formulating their assumptions about the future. The development of probabilistic projections in recent years has triggered an increasing interest in formal methods of expert elicitation.

In 2013, Statistics Canada adopted a rather formal approach to expert elicitation by asking experts to provide estimates of ‘most likely’ values for a series of demographic indicators, along with corresponding 80% prediction intervals. Motivated in large part by a will to improve the characterization of uncertainty, Statistics Canada has recently refined its consultation process, delving further into the science of Expert knowledge elicitation.

In this paper, we describe the expert elicitation protocol used by Statistics Canada in 2018 to inform the development of assumptions. This information may be useful for demographers looking to adopt a formal approach to eliciting expert judgements, and can be pertinent in the context of producing probabilistic projections, where it is necessary – but often difficult – to obtain plausible estimates of uncertainty for components of population growth.

## I. Introduction

1. Expert elicitation can be used to obtain probabilistic information, but with comparatively fewer data requirements (Hanea et al. 2018)—an appealing trait when data are missing or incomplete (Lutz 1994, NRC 2000, Billari et al. 2012, 2014). Time series methods can aptly forecast developments in the future, assuming a continuation in the way that things have evolved in the past (Hyndman and Athanasopoulos 2018). Experts, on the other hand, can envision previously-unseen developments that can steer the future in a new direction, and can take into consideration theories and knowledge from relevant disciplines (Lutz 2009). The ongoing development of probabilistic projections in recent years has triggered new interest in formal methods of expert elicitation, given that these methods offer one way of capturing uncertainty about the future (NRC 2000).
2. Most national statistical offices undertake some kind of consultation with experts when designing their population projection assumptions (UNECE 2018). The scope and format of these consultations vary considerably, ranging from a simple approval procedure from senior management within the organization to the creation of a formal committee of external experts who participate actively in the development of assumptions and methods.
3. In 2013, Statistics Canada conducted a pilot exercise in formal expert consultation to inform its population projection assumption-building process (Bohnert 2015). More recently, Statistics Canada refined its consultation process, designing an elicitation protocol which asks experts to provide complete probability distributions representing a plausible range of the future values of fertility, mortality and immigration in the future. In designing this elicitation protocol, we delved further into the science of *expert knowledge elicitation*, implementing best practices in this regard.<sup>1</sup>
4. The benefits of this new elicitation protocol are numerous, including what we believe to be an improved elicitation experience for the survey respondents, improved accuracy and communication of expert judgments and resulting response aggregation, and more coherent expressions of uncertainty. The latter benefit in particular lends itself well to the direct incorporation of expert judgments into the assumption-building process in both deterministic and probabilistic population projections.
5. In the remainder of this paper, we describe the innovative expert elicitation protocol used in the development of Statistics Canada’s 2018-based population projections (Statistics Canada 2019a, 2019b). Selected results from the protocol are provided, as well as a description of how the results were utilized directly in the building of deterministic projection assumptions. We follow with an application demonstrating how the results from the elicitation protocol could be used in the context of probabilistic projections. We end with some reflections on the utility of this protocol in the further development of probabilistic population projections.

---

<sup>1</sup> There has been much research completed on the challenges associated with expert elicitation. There have also been numerous studies completed on the best methods to counter or minimize those challenges. Readers can find comprehensive reviews of these topics in Garthwaite et al (2005), O’Hagan et al (2006) and Dias et al (2018).

## II. The 2018 Survey of Experts on Future Demographic Trends: Expert Elicitation Protocol

### A. Objectives

6. There are a number of practical criteria that we wanted our elicitation protocol to meet: a small respondent burden (estimated to one hour of work or less), relative simplicity (requiring no extensive expertise in statistics or specialized software knowledge), and low cost of implementation (including the possibility of using remote elicitation). To meet these requirements, it was determined that the design of a MS Excel spreadsheet-based tool offered numerous benefits: the software is widely used, has the ability to incorporate a graphical user interface, and accepts both textual and numerical inputs.
7. A key goal of the protocol, and one sometimes in conflict with the previously-mentioned objectives, was to capture the true belief of the respondent to the greatest extent possible. As part of this objective, *accuracy in the expression of uncertainty* became a main focus of the protocol design. We achieved this by eliciting *complete probability distributions* from experts which, in contrast to eliciting a single point estimate, allows for the expression of the uncertainty about the parameter of interest (Morris et al. 2014). We built our protocol around recent methodological innovations by Keelin (2016, 2018) that led to the development of the *Metalog* distribution; a flexible probability distribution that can be used to model a wide range of density functions using only a small number of parameters elicited from experts. The most appealing feature of this distribution is that it is flexible enough to *accommodate different types of distributions* (for instance, left- or right skewed, bounded or unbounded). We thus avoid making strong assumptions about the characteristics of experts' distributions (e.g., shape, symmetry), and are able to capture nuanced future possibilities.
8. Another way to improve the likelihood of accurately capturing the views of experts is to offer them *visual feedback* associated with their quantitative judgments (Garthwaite et al. 2005, Kynn 2008, Speirs-Bridge et al. 2010, Morgan 2013, Goldstein and Rothschild 2014). In particular, a graphical interface may be more apt to capture people's intuitions about a probability distribution or when otherwise eliciting parameters that are not easy to think about (Jones and Johnson 2014). Visual feedback also allows the respondent to assess, confirm or revise their judgments if desired, thus improving their calibration and accuracy.
9. After eliciting the views of numerous experts, it is necessary to combine their views in some manner. Our protocol's emphasis on the elicitation of complete probability distributions was also driven by the desire to *facilitate the aggregation of experts' responses*, something that is much more difficult and requires many more assumptions when only certain values or quantiles are elicited from experts.
10. These principal objectives, combined with our current knowledge of best practices in elicitation, guided the design of the 2018 expert elicitation protocol, described in the following section.

### B. Design

11. The *2018 Survey of Experts on Future Demographic Trends* was inspired by and builds upon several existing protocols, such as SHELF (Oakley and O'Hagan 2014), EXPLICIT (Grigore et al. 2017), and the self-administered tools designed by Speirs-Bridge (2009) and Sperber et al. (2013) adapted to the remote collection of information from a group of experts.
12. Experts are first presented a short introduction that explains the context and goals of the exercise. They are invited to answer only sections related to components in which they feel they have a certain expertise. Following the introduction, a first set of questions aims at

gathering background information on the respondent. This information is collected for two purposes: firstly, to assess whether the group of respondents is suitably diverse (as recommended by Morgan and Henrion (1990), among others); and secondly, the information is used for the purpose of weighting responses during aggregation, described in more detail in section 2d.

13. The main part of the survey consists of the elicitation of qualitative arguments and quantitative estimates regarding fertility (period total fertility rate), mortality (life expectancy at birth for males and for females) and immigration (number of immigrants per thousand population) for Canada in 2043. The year 2043 was chosen as the target year since it represented the final year in the eventual projection of the provinces and territories. Having a target year 25 years in the future was also deemed to be a good point of balance, forcing experts to think past the short-term evolutions which are likely to follow recent trends, but not so far into the future as to be inconceivable (i.e. we do not ask experts to predict the major demographic behaviours of generations not yet born at the time of the survey). We describe the process using the fertility component as an example.
14. In **Step 1**, we ask for qualitative arguments that are likely to influence the future path of the period total fertility rate in Canada between now and 2043. Experts are also provided a series of tables and figures showing historical trends for various fertility indicators. Experts are invited to think about a variety of possible future scenarios (increase, decrease, status quo) when formulating their arguments. Besides providing critical information for putting into context their later quantitative estimates, this procedure is recommended as it encourages experts to think about the substantive details of their judgments and consider a whole range of possibilities, thus reducing potential overconfidence (Morgan and Henrion 1990, Kadane and Wolfson 1998, Garthwaite et al. 2005, Kynn 2008).
15. **Step 2** is modelled in large part by the step-based procedures utilized by Speirs-Bridge et al. (2010), Sperber et al. (2013) and Grigore et al. (2017) and comprises four subparts:
  - a) Experts are first asked to provide the lower and higher bounds of a range covering nearly all plausible<sup>2</sup> values of the period total fertility rate in Canada in 2043. Beginning with the contemplation of the extremes of the distribution is an intentional practice used to minimize potential overconfidence (Speirs-Bridge et al. 2010, Sperber et al. 2013, Oakley and O'Hagan 2014, Grigore et al. 2017, Hanea et al. 2018). Indeed, asking experts to first provide a single central estimate such as a mean or a median tends to could trigger anchoring to that value in subsequent responses.
  - b) Experts are asked to report how confident they are that the true value will fall within the range they just specified in step 2(a). Allowing experts to determine their own level of confidence has been found to reduce overconfidence in comparison with asking them to identify the low and high bounds of an interval to some predetermined confidence level (Speirs-Bridget et al. 2010).<sup>3</sup>
  - c) Experts are asked to estimate the median value of the plausible range they provided in step 2(a), so that they expect an equal (50-50) chance that the true value lies above or below the median.
  - d) The range of values between the lower bound and the median is split in two segments of equal length and the same is done for values between the median and the upper bound.

---

<sup>2</sup> The term “plausible” was arrived at after much careful consideration. As illustrated by Morgan (2013), terms such as “probable”, “likely”, or “possible” may be interpreted very differently by different respondents.

<sup>3</sup> That said, we impose the restriction that the respondent must choose a confidence level of at least 90% or higher; experts are asked to revise their range if they are confident at a level of less than 90%.

The respondent is then asked to assign to each segment the probability that the true value falls within each of these segments. Note that each half below and above the median has by definition 50% probability of occurrence, so it is a matter of redistributing that 50% to each segment.<sup>4</sup>

16. Throughout step 2, several “checks”, in the form of pop-up warning signs, were built into the elicitation tool in order to prevent illogical inputs in various forms.
17. We used Keelin’s *Metalog* distribution (2016, 2018) to calculate each experts’ probability density function based on their responses to the questions above. The metalog distribution – short for “meta-logistic” – belongs to the larger class of Quantile-Parameterized Distributions (QPDs) developed by Keelin and Powley (2011), and refers to any continuous probability distribution that can be fully parameterized in terms of its quantiles. The appeal of using QPDs in modelling uncertainty is that modifications can be made to their quantile functions (through the addition of extra shape parameters, for example), enabling them to represent a broader range of beliefs.
18. The “meta” in metalog is a term used by Keelin to describe distributions whose original parameters have been substituted in order to incorporate a greater number of shape parameters. In theory, there is no limit to the number of shape parameters the metalog distribution can have, meaning it can be used to model distributional characteristics such as right- or left-skewness, varying levels of kurtosis, and multi-modality. Since the parameters of the metalog are a function of its quantiles, however, the inclusion of additional shape parameters requires the elicitation of a greater number of quantiles. The procedure described in step 2 is designed to elicit five quantiles, enabling the algorithm to fit unbounded metalog distributions with up to a maximum of five shape parameters. In the event that experts’ inputs describe a semi-bounded or bounded distribution, log- or logit-transforms are applied to the metalog quantile function, respectively, in order to restrict its range accordingly.
19. Moving next to a key and innovative feature of our protocol: in **step 3**, respondents are provided with a visual representation of the parameter estimates they provided in step 2, in the form of a histogram and probability density function). Although we chose to elicit values that are most easily understandable (i.e. median and probabilities instead of parameters of parametric distributions such as mean and variance), it may not be easy for an expert to grasp how a change in median value will precisely influence the corresponding probability distribution. As mentioned earlier, visual feedback allows experts to test if their inputs generate a result corresponding to what they had in mind and reconsider their estimates if desired (Kynn 2008). Implementation of the visual interface was relatively easy thanks to Keelin’s free MS Excel distribution program.
20. Despite being highly flexible, there can be instances where our version of the metalog algorithm (having a maximum of 5 shape parameters) is unable to compute a probability density function given the inputs provided. This can occur for example if an expert envisions a largely bimodal probability density function. For this reason, a rudimentary histogram is also presented to the expert which, despite not accurately representing the tails of their envisioned distribution, still reflects their inputs in a crude manner, allowing them to recognize any possible mistakes they may have made or possible biases they may have been

---

<sup>4</sup> This represents the fixed interval method. We found in testing that it performed better than the variable interval method in minimizing the range-principle effect (see Parducci 1963), a problem that has been reported in other elicitation exercises (e.g., Sperber et al. 2013, Gosling 2014). In comparison with the variable interval method, respondents found the task easier and more intuitive with the fixed interval method, and their responses were more plausible.

subjected to. When a probability density function cannot be computed, experts are informed and instructed to go to the next step if they nevertheless feel comfortable with their inputs.<sup>5</sup>

21. Once experts have reviewed the graphed densities and are satisfied with their inputs, they are invited to comment on the results in step 4. They are also asked to indicate to what extent the resulting probability density function represents an accurate description of their beliefs (i.e. very accurate, good, poor). Lastly, experts who answered that the visualization of the results did not provide a coherent representation of their beliefs are asked to provide further explanation.

### C. Survey results

22. Members of Canada's two demography associations, the Canadian Population Society and l'Association des démographes du Québec, were invited to complete the *2018 Survey of Experts on Future Demographic Trends* questionnaire remotely. In the context of an elicitation on the topic of Canadian demography—a very small field of academic discipline, narrowed further by the fact that we were asking specifically about the future, requiring some level of familiarity with demographic projections—experts are a fairly scarce resource. In total we received 18 responses to the survey. Respondents were found to represent a fairly well-balanced mix of expertise, general years of experience in the field, and current domain of work. The majority of respondents (10 out of 18) reported having high levels of expertise in demographic projections. By and large, respondents reporting low or no expertise in a given component elected to skip the questions relating to that component, as was expected.

### D. Aggregation of individual responses

23. After eliciting the views of numerous experts, it is necessary to combine their views in some manner. The choice of aggregation method was made with the goal of capturing as much information as possible from the experts' individual beliefs, while ensuring that the aggregate result is itself a valid probability distribution from which relevant summary statistics—such as the mean, median, and quantiles—can be derived. For this reason, we adopted a mixture model approach (referred to as a “linear opinion pool” when applied to the context of expert elicitation) in which the aggregate distribution for each component can be thought of as a weighted average of the individual expert distributions. Linear pooling is simple, transparent, and in comparison to other methods, tends to yield distributions with more dispersion, thus offsetting the effect of experts' overconfidence, if present
24. Each expert's contribution was weighted on the basis of their self-assessed level of experience about the different components of growth and in population projections. We preferred to weight responses in the context where we solicit a large number of experts in demography with varying levels of expertise in the areas of fertility, mortality, immigration. It also seemed appropriate in the case where a respondent reports a low level of expertise in a given demographic component and somehow expects us to take this information into account.
25. Despite the fact that experts' responses are parametrized by metalog distributions, the resulting mixture distributions for fertility, mortality, and immigration are not metalog distributions, and do not belong to any defined parametric family. Characteristics such as central moments and quantiles are derived using numerical methods. **Figure 1** illustrates the

---

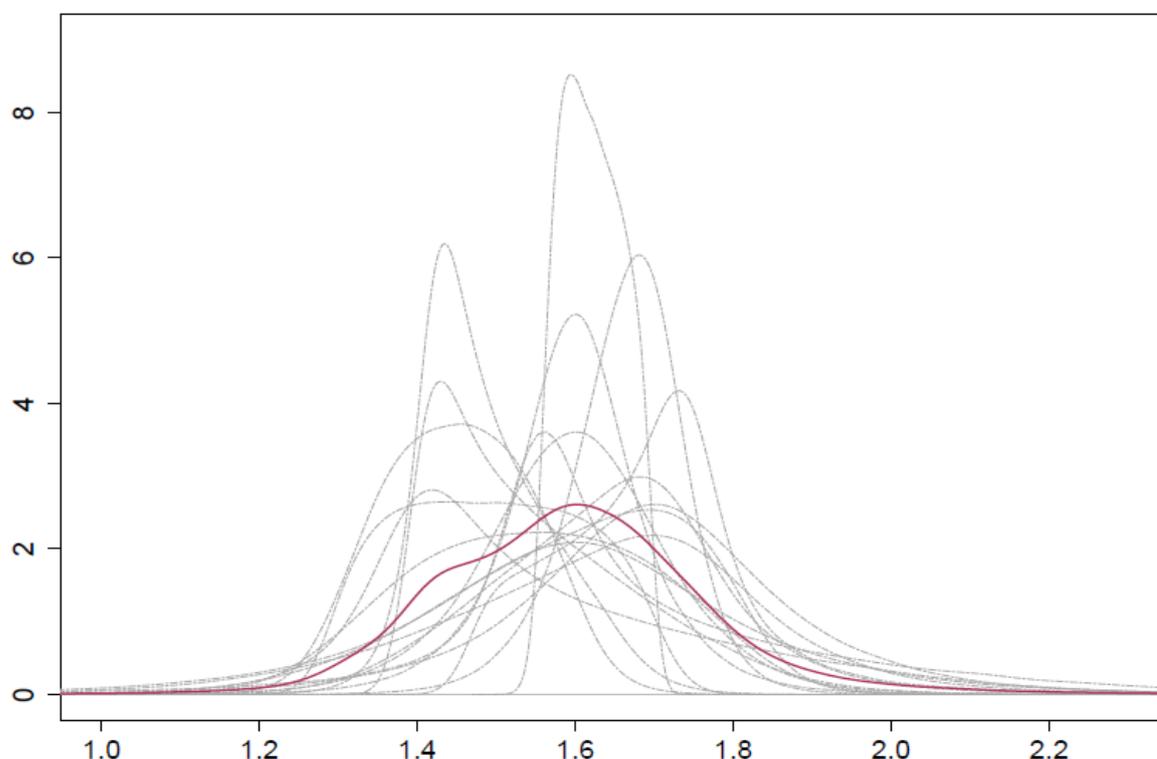
<sup>5</sup> The idea is that since an infinite number of distributions could correspond to their inputs, their inputs may be faithful to their assessments of the future, even though a visual representation could not be produced. The histogram remains useful as a way to validate their inputs.

individual probability distributions provided by experts regarding the plausible range of the period total fertility rate in Canada in 2043 and resulting aggregate mixture distribution.

### E. Incorporation of expert judgments into the deterministic projection assumption-building process

26. The aggregate mixture distributions described in the preceding section represent experts' views in 2043, but values are also needed for all interim years of the projection. As Lee (1998) rightly pointed out, expert opinion may be of little help for forecasting intermediate years without information about the autocorrelation structure. This is why we make no inference about what experts had in mind regarding the interim evolution leading to the 2043 distribution; instead, we make our own assumptions about it. To make these assumptions, we privileged time series models, for their capacity to provide probabilistic development over time informed by historical data, calibrated to match experts' densities in 2043. The rationale for this 'hybrid' methodology is that while experts can go beyond past trends and include more information in thinking in the long term, time series models can aptly forecast future trends replicating past autocorrelations—information that experts would have difficulty envisioning. We therefore see this approach as a balanced mix of utilization of time series modelling and expert opinion, benefitting from each method's strengths.

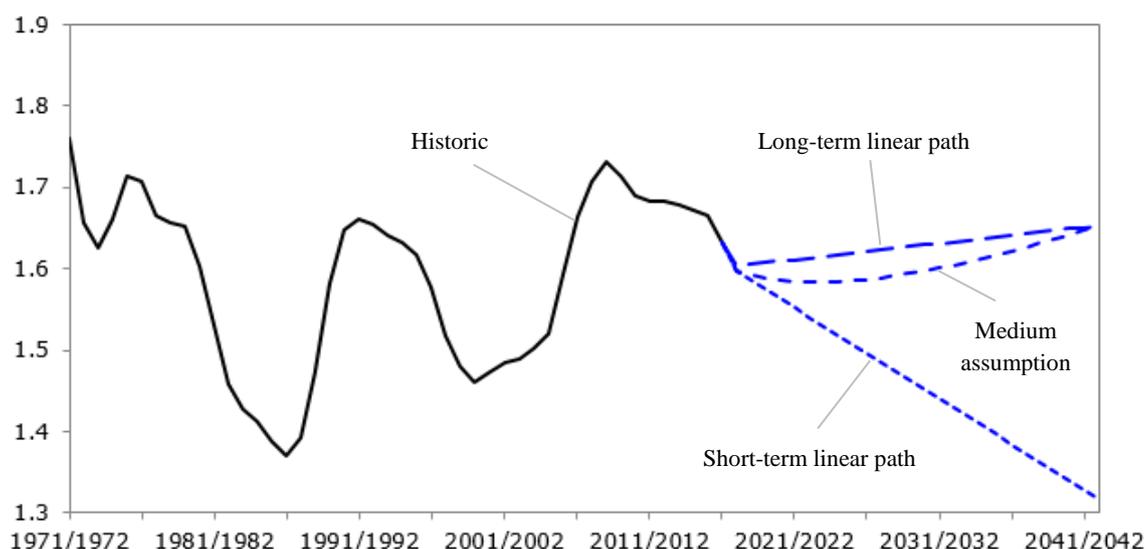
**Figure 1** Period total fertility rate, Canada, 2043: Individual expert probability distributions (grey dashed curves) and aggregate mixture distribution (red curve) of the 17 fertility respondents of *the 2018 Survey of Experts on Future Demographic Trends*



Source: Statistics Canada, Demography Division.

27. The targets obtained from the survey at the Canada-level are used to derive the regional targets, assuming the same proportional growth in percentage. This method is consistent with the traditional “hybrid bottom-up” approach often used in population projections: assumptions specific to each region are constructed from assumptions initially developed at the national level, but the Canada-level projections exist only by summing the results for the provinces and territories individually. Briefly, medium assumptions for each component are derived as follows:
- Two distinct linear trajectories are produced for the period 2018 to 2043 for each of the provinces and territories: (1) a short-term trajectory based on the examination of historical trends, and (2) a long-term trajectory based on the results from the *2018 Survey of Experts on Future Demographic Trends*. The 50th percentile (median) of the aggregate expert distribution was used as the long-term national target in 2043.
  - These two linear trajectories are combined to obtain a single medium assumption, with the use of a logarithmic interpolation technique that allows for a smooth transition.
28. The logarithmic interpolation of the two short- and long-term trajectories, yielding a single assumption, makes use of weights selected so that the curve based on the short-term trajectory is given more weight earlier on in the projection years, and the curve based on the long-term trajectory is given more weight in the latter years. The consequence is that in the short-term, assumptions for a given province will reflect mostly recently observed trends, whereas in the long-term, they will be more influenced by beliefs about future trends at the Canada level. Using logarithmic interpolation (as opposed to linear interpolation, for instance), ensures that the short-term trajectory fades relatively quickly in favour of the long-term trajectory. This approach follows best practices in projections to consider the plausibility of outcomes for multiple horizons, in contrast to focusing solely on long-term outcomes (UNECE 2018). **Figure 2** provides an example of the projected period total fertility rate in the province of Québec according to the medium assumption. The graph displays the short-term trajectory, long-term trajectory and final medium assumption. More details about the methodology can be found in Statistics Canada (2019b).

**Figure 2** Period total fertility rate, Quebec, historic (1971/1972 to 2016/2017) and projected (2017/2018 to 2042/2043)



**Note:** The 2017 data are considered preliminary.

**Sources:** Statistics Canada, Canadian Vital Statistics, Births Database, 1971 to 2017, Survey 3231 and Demography Division.

29. Low and high assumptions were built based on the medium assumption described above, with targets reflecting experts' uncertainty. The low assumption long-term target (for 2043), was computed by taking the 10th percentile of the aggregate probability distribution of experts, and the high long-term target was computed by taking the 90th percentile. Thus, low and high long-term targets represent the bounds of an 80% prediction interval around the medium long-term target. Again, for brevity, we refer readers to Statistics Canada 2019b for more information about the methodology.

### III. Application: using the *2018 Survey of Experts on Future Demographic Trends* to produce probabilistic projections of the PTFR

30. Producing probabilistic projections of the population requires obtaining probabilistic information on the individual components of population growth. The primary difficulty associated with this is correctly identifying both the individual autocorrelation structures of each of the components, and the structure of the temporal cross-correlation between components. This task becomes exceedingly complex when projections at the subnational level are desired, as regional correlations must also be considered.
31. In this section, we expand on a method developed by Lutz et al. (2001) in order to provide an example of how results from the *2018 Survey of Experts on Future Demographic Trends* can be combined with traditional time series models – which provide an autocorrelation structure – to produce probabilistic projections. For simplicity, we limit ourselves to the projection of a single demographic indicator (the PTFR) at the national level.<sup>6</sup>

#### A. Method

32. The method utilizes ARIMA models in combination with a priori knowledge about certain properties of the forecast distribution of a given component to derive the full forecast distribution. More specifically, the method assumes that the forecast variance of a component in some year  $t$  of the forecast is known, and that the time series parameter(s) can be selected in such a way that the target variance is met in the desired amount of years.<sup>7</sup> Briefly, the model (of, for e.g. the PTFR) can be represented in the following way:

$$y_t = \bar{y}_t + \varepsilon_t$$

$$t = 1, \dots, T$$

33. Where  $y_t$  represents the value of the PTFR in year  $t$  of the projection,  $\bar{y}_t$  represents the mean value of the PTFR in year  $t$  (also assumed to be known in advance), and  $\varepsilon_t$  represents the deviation from the mean value in year  $t$ . The standard deviation of the error in year  $t$ ,  $\sigma(\varepsilon_t) = \sigma(y_t)$ , is predetermined according to assumptions about the expected level of

---

<sup>6</sup> This method can be extended further to allow for the projection of multiple components and regions, including estimating cross-component and regional correlations. See the supplementary material provided in Lutz et al. (2001) for details.

<sup>7</sup> A full description of this method is provided in the supplementary material of Lutz et al. (2001).

future projection uncertainty. In Lutz et al. (2001), a combination of expert opinion and the ex-post analysis of past projection errors is used to obtain standard deviation targets. Given this information, a moving average model of order  $q$  (MA( $q$ )) is used to model the  $\varepsilon_t$ , with the parameters of the model selected in such a way that  $\sigma(\varepsilon_t)$  is equal to its pre-specified target.<sup>8</sup> To generate prediction intervals, 2,000 simulations are produced.

34. We modify this method in order to incorporate the aggregate expert probability distribution of the PTFR in 2043 obtained from the survey. Similar to Lutz et al., we use a MA( $q$ ) model with  $q = 26$ <sup>9</sup> to produce projections of the PTFR from 2018 to 2043, with additional calibration parameters to ensure that in the last year of the projection (in 2043, or when  $t = 26$ ), the forecast distribution obtained from the time series model is identical to the one obtained from the survey. The method is summarized below.
- 1) The mean of the survey distribution is used as the target mean in 2043; i.e.,  $\bar{y}_{26} = \bar{y}_{survey}$ . Target means for the intermediary years,  $t = 1, \dots, 25$ , are obtained using the logarithmic interpolation technique used to derive the medium projection assumption for the PTFR outlined in the previous section. The mean series  $\bar{y}_t$  thus reflects both recently observed trends in the PTFR and beliefs about its future long-term level obtained from the survey.
  - 2) The target standard deviation of the error in year  $t = 26$  – also the standard deviation of the forecast,  $\sigma(y_{26})$  – is set to the standard deviation of the survey distribution.<sup>10</sup> The 27<sup>11</sup> moving average parameters in the MA(26) model are then set to equal  $\sigma(y_{26})/\sqrt{27}$ , which guarantees the standard deviation at  $t = 26$  is equal to its target value. This also determines the standard deviation in years  $t = 1, \dots, 25$ , as they are a function of the moving average parameters.
  - 3) Once the mean and standard deviation targets are selected, the full forecast distribution is then obtained using the following 5-step algorithm:
    - a) 100,000 values from the expert survey distribution are drawn at random, and then ranked. The empirical mean,  $\bar{y}_{survey}$ , and standard deviation,  $\sigma(y_{survey})$ , are computed.
    - b) 100,000 simulations from an MA(26) model are produced for years 2018-2043, with the forecast mean series selected as in 1) and the forecast variance as in 2). Simulations are then ranked in terms of their value in the last year, 2043.

<sup>8</sup> Parameters are not estimated using historical data as is normally the case with time series model. Instead, parameters are derived analytically, conditional on some known properties of the forecast distribution (i.e. the variance).

<sup>9</sup> The choice of  $q$  depends on the length of the projection period, as well as what point in the period the desired variance target should be met. An MA( $q$ ) model is typically forecastable a maximum of  $q$ -periods-ahead.

<sup>10</sup> Setting the standard deviation target to the standard deviation of the expert distribution before calibration guarantees that post-calibration, the structure of the forecast variance remains unchanged. For example, an unmodified MA(26) model with parameters specified as in Lutz et al. (2001) has a forecast variance that increases linearly throughout the projection. By setting the standard deviation of the MA(26) model to the survey standard deviation at  $t = 26$ , even after calibration parameters have been added to shift the forecast distribution over the course of the projection, the forecast variances in each year remain unchanged (i.e. they increase linearly).

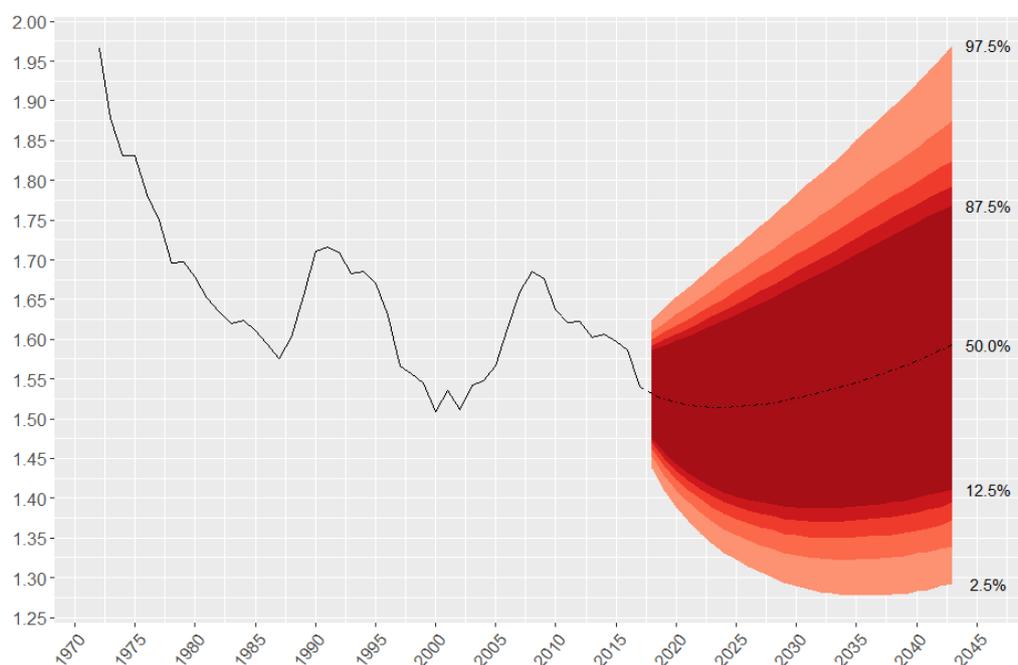
<sup>11</sup> Unlike in a standard MA( $q$ ) model, the first term in the moving average series as specified in Lutz et al. (2001) does not have a parameter of 1.

- c) Each ranked simulation is paired with its corresponding ranked draw from the mixture distribution; i.e. the fifth draw is paired with the fifth simulation.
- d) The difference between the simulation value in 2043 and its paired draw is computed, and a constant is added to each simulation so that in 2043, the values are the same. The constant is added proportionally over the course of the simulation so that the calibration procedure doesn't cause a "shock" that shifts the simulation drastically.<sup>12</sup>
- e) The empirical distribution of the time series forecast at year 2043 is now identical to that of the survey distribution, with the mean series  $\bar{y}_t$  remaining unchanged. Percentiles can be computed empirically in order to obtain prediction intervals about the median of the forecast distribution.

## B. Results

35. **Figure 3** displays select percentiles of the forecast distribution of the PTFR from 2018-2043, along with the historical series (1972-2017). The dotted line (50th percentile) is consistent with the Canada-level medium assumption for the PTFR in *Population Projections for Canada (2018 to 2068), Provinces and Territories (2018 to 2043)* (Statistics Canada, 2019a).<sup>13</sup>

**Figure 3** Historical period total fertility rate (1972-2017) and select percentiles of the forecast distribution (2018-2043)



**Note:** The dotted black line corresponds closely to the Canada-level medium projection assumption for the PTFR in *Population Projections for Canada (2018 to 2068), Provinces and Territories (2018 to 2043)*.

**Source:** Statistics Canada, Canadian Vital Statistics, Births Database, 1977 to 2017, Survey 3231 and Demography Division.

<sup>12</sup> i.e. a constant is added to the value at every year in the simulation, and not simply to the value in 2043.

<sup>13</sup> A more detailed description of the projections methodology can be found in *Population Projections for Canada (2018 to 2068), Provinces and Territories (2018 to 2043): Technical Report on Methodology and Assumptions* (Statistics Canada, 2019b).

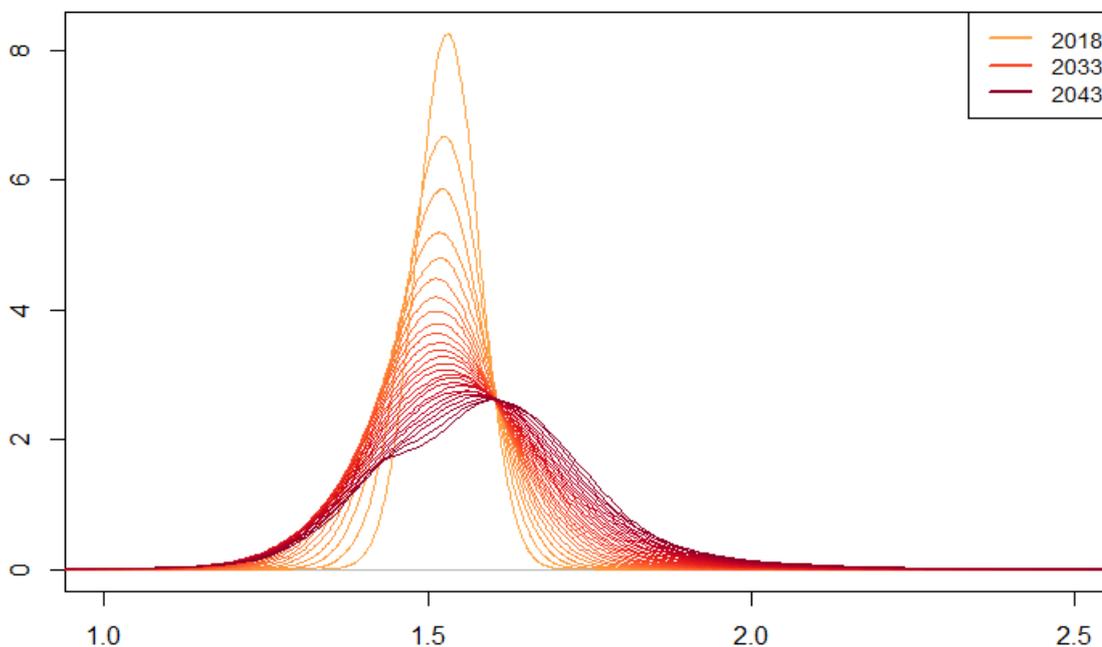
36. It should be noted that while the forecast distribution in 2043 is determined by the survey distribution, the forecast distribution in all other years of the projection is determined by the selected parameters ( $\bar{y}_t$  and  $\sigma(y_t)$ ) as well as the initial forecast distribution of the  $\varepsilon_t$  terms. Given that the  $\varepsilon_t$  series is modelled by an MA(26), a single simulation from this model can be parameterized in the following way:

$$\varepsilon_t = \sum_{i=0}^{27} \alpha_i u_{t-i}$$

$$u_{t-i} \sim iid N(0,1)$$

37. Where  $\alpha_i$  are the moving average parameters. Thus, prior to calibration,  $y_t \sim N(\bar{y}_t, \sigma(\varepsilon_t))$ . By adding a constant to each simulation to shift the forecast distribution in 2043, the forecast distribution in all other years gradually shifts from being Normal (or approximately Normal) in earlier years of the projection toward a distribution more similar to the metalog mixture in later years of the projection.
38. **Figure 4** shows the evolution of the forecast density over the course of the projection. The lighter, orange lines display the distribution in earlier years (symmetrical and approximately Normal) and the darker red lines display the distribution in later years (asymmetrical and closer in shape to the survey mixture metalog distribution). The darkest line, representing the distribution in 2043, is that of the expert distribution. The unusual shape of this distribution suggests that traditional time series models that impose a Normal forecast distribution across all years would fail to accurately represent the aggregate information conveyed by experts.

**Figure 4** Forecasted density of the period total fertility rate, 2018-2043



Source: Statistics Canada, Demography Division

## C. Future developments

39. This method of producing probabilistic projections can be thought of as a simulation-based approach that makes minimal assumptions about the autocorrelation structure of the process. Given the only information known about the full forecast distribution prior to producing projections is: (1) the mean at every year in the projection; and (2) the distribution in the last year of the projection, deriving conditional distributions at all other years requires making no small number of assumptions about the underlying data generating process. ARIMA models, or variations of them, have long been utilized in the projection of fertility (see for example Lee and Tuljapurkar, 1994, Keilman and Pham, 2004, and Alders and de Beer, 2004) as well as other demographic indicators. Using simulations from an MA model as a starting point provides both a plausible correlation structure and an initial distributional assumption (Normal).
40. The way these simulations are modified so that the distribution in the last year of the projection reflects the survey distribution rather than the Normal distribution produced by a standard MA model, however, modifies these assumptions indirectly. The addition of a constant to shift the individual simulations modifies the conditional densities gradually over time, while maintaining the same mean and variance.<sup>14</sup> This process is equivalent to simulating values from the chosen model without explicitly formulating it; the final model is not an MA, but its true form is not derivable – nor does it need to be – from the modified simulations.
41. In practice, any type of ARIMA model can be used to generate probabilistic projections using this approach. Lutz et al. (2001) use both AR and MA models to generate probabilistic forecasts, and argue that the choice between either model should be based on which appears to better accurately reflect the future. Our modified approach is largely insensitive to the choice of initial model due to the modification process.<sup>15</sup> Assuming a Normal distribution at the start of the projection and the expert distribution at the end restricts the number of ways the process can evolve over time.
42. The most difficult aspect of such an approach remains combining it for a number of different indicators (e.g. life expectancy and migration) and across different regions. It is likely that a number of simplifying assumptions will need to be made in order to estimate correlations between both components and regions – in the literature, for example, it is sometimes assumed that components are independent, or that correlations are insignificant enough to be ignored (Billari et al., 2012, Keilman, 1997, Lee and Tuljapurkar, 1994). Estimates of correlation may also be elicited formally through expert opinion (e.g., Billari et al., *ibid.*), though this comes at the cost of significantly increasing the burden on respondents. The approach developed by Lutz et al. (2001) that we adapted in this section does allow for the production of probabilistic projections correlated across different regions, or across distinct components of population growth. However, we leave these considerations and the potential extension of this model to future work.

---

<sup>14</sup> An attractive feature of obtaining a normal distribution at the start of the projections is that it is the distribution that makes the least assumptions (i.e. admits the most ignorance) beyond what is stated, here, a known mean and standard deviation (the standard deviation resulting from the chosen MA process). In this context, the normal distribution is the one with the largest entropy. The distribution changes over time as we approach the year 2043, for which we assume having full knowledge.

<sup>15</sup> The approach has only been tested with AR, MA, and random walk (RW) models. Whether this is true for other specifications has not yet been determined.

## IV. Conclusion

43. We used expert elicitation as a way to better inform the assumption-building process of deterministic scenario-based projections. The resulting scenarios have interesting properties: they share the same definition from one component of growth to another, and they are anchored in real probabilistic information coming from the experts and past data. One of the key advantages of this new approach to projection assumption-building is its conceptual consistency across components: the long-term projection assumptions share the same probabilistic meaning: the “high” assumption represents the 90th percentile of the aggregate probability distribution of plausible future values for that given component according to the experts who responded to the survey; the “medium” assumption represents the 50th percentile, and the “low” assumption the 10th percentile. This leads to greater coherence in the resulting projection scenarios (which combine assumptions about the various components).
44. Looking forward, the elicitation protocol described in this article can be used to produce a large number of stochastic trajectories that could be combined for the production of probabilistic projections, either as described in the previous section or by utilizing alternative methods.

---

## References

- Alders, M. and de Beer, J. (2004). Assumptions on fertility in stochastic population forecasts. *International Statistical Review* 72(1): 65-79
- Billari, F.C., Graziani R., and Melilli, E. (2012). Stochastic population forecasts based on conditional expert opinions. *Journal of the Royal Statistical Society A* 175(2): 491–511.
- Billari, F.C., Graziani R., and Melilli, E. (2014). Stochastic population forecasting based on combinations of expert evaluation within the Bayesian paradigm. *Demography* 51:1933–1954.
- Bohnert, N. (2015). Chapter 2: Opinion Survey on Future Demographic Trends. In: Bohnert, N., Chagnon, J., Coulombe, S. Dion, P. and Martel, L. *Population Projections for Canada (2013 to 2063), Provinces and Territories (2013 to 2038): Technical Report on Methodology and Assumptions*. Statistics Canada catalogue number 91-620-X.
- Dias L., Morton, A., and Quigley, J. (eds). 2018. *Elicitation: the science and art of structuring judgment*. Springer, New York.
- Garthwaite, P.H., Kadane, J.B., and O’Hagan, A. (2005). Statistical methods for eliciting probability distributions. Technical paper 1-2005. Carnegie Mellon University Research Showcase.
- Goldstein, D.G. and Rothschild, D. (2014). Lay understanding of probability distributions. *Judgment and Decision Making* 9(1): 1-14.
- Gosling, J.P. (2014). Methods for eliciting expert opinion to inform health technology assessment, Vignette on SEJ methods for MRC, UK, 2014.
- Grigore, B. Peters, J., Hyde, C. and Stein, K. (2017). EXPLICIT : a feasibility study of remote expert elicitation in health technology assessment. *BMC Medical Informatics and Decision Making* 17: 131.
- Hanea, A. M., Burgman, M., and Hemming, V. (2018). Chapter 5: IDEA for Uncertainty Quantification. In Dias L., A. Morton, J. Quigley (eds) *Elicitation: the science and art of structuring judgment*. Springer, New York.
- Hyndman R.J. and Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*, Monash University, Australia. Online version at <https://otexts.org/fpp2/>
- Jones, G. and Johnson, W.O. (2014). Prior elicitation: Interactive spreadsheet graphics with sliders can be fun, and informative. *The American Statistician*, 68(1): 42-51.
- Kadane, J.B. and Wolfson, L.J. (1998). Experiences in Elicitation. *Journal of the Royal Statistical Society. Series D* 47(1): 3-19.
- Keelin, T.W. 2016. The Metalog Distributions. *Decision Analysis* 13(4).

- Keelin, T.W. 2018. The Metalog Distributions – Excel workbook. Accessible at: <http://www.metalogdistributions.com/excelworkbooks.html>.
- Keelin, T.W. and Powley, B.W. 2011. Quantile-parameterized distributions. *Decision Analysis* 8(3): 165-250.
- Keilman, N. (1997). Ex-post errors in official population forecasts in industrialized countries. *Journal of Official Statistics* 13: 245-277.
- Keilman, N. and Pham, D.Q. (2004). Time series based errors and empirical errors in fertility forecasts in the Nordic countries. *International Statistical Review* 72: 5-18
- Kynn, M. (2008). The ‘heuristics and biases’ bias in expert elicitation. *Journal of the Royal Statistical Society, Series A (Statistics in Society)* 171(1): 239-264.
- Lee, R.D. and Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: Beyond high, medium, and low. *Journal of the American Statistical Association* 89: 1175-1189
- Lee, R. (1998). Probabilistic approaches to population forecasting. *Population and Development Review* 24: 156-190.
- Lutz, W. (ed.). (1994). *The Future Population of the World: What Can We Assume Today?* London: Earthscan.
- Lutz, W., Sanderson, W.C. and Scherbov, S. (1998). Expert-based probabilistic population projections. *Population and Development Review* 24:139-155
- Lutz, W., Sanderson, W.C. and Scherbov, S. (2001). The end of world population growth. *Nature* 412: 543-545
- Lutz, W. (2009). Toward a Systematic, Argument-Based Approach to Defining Assumptions for Population Projections. Interim Report IR-09-037. Laxenburg, International Institute for Applied Systems Analysis.
- Morgan, M, G.. and Henrion, M. (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press.
- Morgan, M.G. (2013). Use (and abuse) of expert elicitation in support of decision making for public policy. *PNAS Perspective* 111(20): 7176-7184.
- Morris, D.E., Oakley, J.E. and Crowe, J.A. (2014). A web-based tool for eliciting probability distributions from experts. *Environmental Modelling and Software* 52: 1-4.
- NRC. [National Research Council] (2000). *Beyond six billion. Forecasting the world’s population*. Washington, DC: National Academies Press.
- Oakley, T. and O’Hagan, A. (2014). SHELF: the Sheffield elicitation framework (version 3.0). <http://www.tonyohagan.co.uk/shelf/SHELF3.html>. Accessed June 15<sup>th</sup> 2018.
- O’Hagan A, Buck CE, Daneshkhah A, Eiser JR, Garthwaite PH, Jenkinson DJ, et al. (2006). *Uncertain Judgements: Eliciting Experts’ Probabilities*. Chichester: Wiley.
- Parducci, A. (1963). Range-frequency compromise in judgment. *Psychological Monographs: General and Applied* 77(2): 1-50.
- Sperber, D., Mortimer, D., Lorgelly, P. and Berlowitz, D. (2013). An expert on every street corner? Methods for eliciting distributions in geographically dispersed opinion pools. *Value in Health* 16: 434-437.
- Speirs-Bridge, A., Fidler, F., McBride, M., Flander, L., Cumming, G., and Burgman, M. (2010). Reducing overconfidence in the interval judgments of experts. *Risk Analysis* 30(3): 512-523.
- Statistics Canada. (2019a). *Population Projections for Canada (2018 to 2068), Provinces and Territories (2018 to 2043)*. Statistics Canada catalogue number 91-520-X.
- Statistics Canada. (2019b). *Population Projections for Canada (2018 to 2068), Provinces and Territories (2018 to 2043): Technical Report on Methodology and Assumptions*. Statistics Canada catalogue number 91-620-X.
- UNECE. (2018). Recommendations on Communicating Population Projections. ECE/CES/STAT/2018/1. Available at: <http://www.unece.org/statistics/networks-of-experts/task-force-on-population-projections.html>.