**1 Title Slide**

Hello, my name is Ricarda Duerst, I am PhD student at the University of Helsinki and the Max Planck Institute for Demographic Research.

I am going to present ongoing work on forecasts of Nordic period fertility, including Finnish fertility, with empirical prediction intervals. My colleagues on this project are all here: Julia, Jessica, Mikko, and my supervisor Jonas Schöley, present online.

I am going to focus on our research plan, the concept of empirical prediction intervals, and in the end, I will show some preliminary results.

**2 TFR**

To start off, a little reminder what the so-called Nordic Fertility Regime looks like. This is the Total Fertility rate over time in Finland, Denmark, Sweden, and Norway. The Finnish data is green. We have high variability during the first half of the 1900s, strong decreases during the 1970s. Followed by less variability under the replacement level with moderate increases until 2010. And then, another drop in TFR to all-time lows, at least for Finland and Norway, during the 2010s until now. I guess I don’t have to mention the consequences of fertility levels that low for social security and pension systems.

**3**

That brings us to the question of what is needed, or what **you** want? Forecasts of Nordic Fertility and sensible prediction intervals. On the other hand, we have what science is providing us with: a colourful bouquet of forecast models and model-based prediction intervals that are usually too narrow.

**4 Research Plan**

To provide you with what you want, we came up with a research plan. … (see slide)

**5 Forecast Models**

Before I come to the empirical prediction intervals, I will show you our selection of forecast models, that we have chosen so far.

**6 PPS**

First of all, there is the postponement scenario. This model was also part of the last report from 2020 and was developed by Jessica, Julia, Pekka Martikainen and Mikko. The main assumption of this scenario-based forecast model is, that the fertility postponement that has verifiably happened and that caused the decrease of the TFR in the last decade, will continue, but slow down. And eventually it will stop. Which means that the tempo-adjusted TFR and the normal TFR without consideration of tempo-effects such as postponement of births, will converge by the end of the forecast horizon. That is why in this forecast from 2020 to 2040, the TFR is increasing and then levelling off.

**7 Lee-Carter**

Second model is the lee-carter model for fertility. This is an extrapolation-type forecast model and a demographic classic. It is based on the same method as the lee-carter model for mortality. Using singular value decomposition, age-specific and time-specific components are estimated and the time-specific components are then forecast using a random walk with drift.

**8 ARCH**

So far, we have chosen a third model, which is more commonly used in economics than in demography. However, Nico Keilman and Pham argue and show, that it is also a nice fit for TFR forecasts. The Auto-regressive conditional heteroskedasticity model, let’s call it ARCH, combines time-varying variance levels (that’s the heteroskedasticity part) with an autoregressive-process. As Keilman and Pham argue, it fits the Nordic TFR, given the varying levels of variability in TFR throughout the time series. We have seen this in the plot: in the first half of the last century, there TFR variability was higher, than in more recent years.

And finally, we also included a naïve forecast approach as reference. We simply fix the last observed value of the TFR. We are open to include additional forecast models, as well.

**9 Empirical Prediction Intervals**

Now, for the exciting part. What are empirical prediction intervals?

**10 Principle of Empirical Prediction Intervals**

The idea or principle behind them is simple: A forecast is only as precise as similar forecast in the past turned out to be. Which means that we need historical forecasts to calculate them. The steps to derive empirical prediction intervals can be divided into steps of the calibration, the validation, and the application. The application in this case are real forecasts for Nordic fertility into the future.

**11 Sleeping Beauty**

The idea of empirical prediction intervals is not a new one. Originating in Economics, Williams and Goodman were probably the first to introduce the concept. And 12 years later they made an appearance in demography and was forgotten again. Ronald Lee, the one of the authors of the Lee-Carter method, puts it like this 1998: “[Empirical prediction intervals are] a very valuable line of research that has been largely neglected since the publication of [the paper by Stoto].” But then in 2004 Nico Keilman and Dinh Pham did something very similar to what we are doping: They applied empirical prediction intervals to fertility forecasts for the Nordic countries. Demography and Economics are however, not the only field of research where they appear. In machine learning the concept is known under the term conformal prediction. And now after more than 50 years since the first publication, it is our turn to revive the empirical prediction intervals.

We are not only using them because a few leading scientists think they are great. But because they have advantages for our case. First of all, history has shown that model-based prediction intervals for population forecasts are too narrow. Second, I think you know this best, you have to be prepared for exceptions and that is why in risk management we need to make probabilistic decisions in the face of risks. Or, as Jonas has put it, you wouldn’t just evacuate the county in which the median of your hurricane forecast lies, but also the surroundings. And history has shown us, that there is a non-neglectable risk to be very wrong with our population forecasts. The error distribution is long-tailed. And therefore, there are qualitative differences between model-based and empirical prediction intervals.

**12 Empirical Prediction Intervals**

Let’s get to business. How do we derive empirical prediction intervals for the Nordic countries? First, we take the available time series of fertility for each country and cut them up into smaller pieces. These are our cross-validation series. Because different data availability the number of series varies between countries. Then each series is divided into a training period, blue, and a testing period, pink. To the training period we apply a forecast method, and the forecast results will then be compared to the observed data in the testing period.

**13**

When we do this for all the cross-validation series, we can plot the forecast error distribution over the forecast length. This can now be either modelled by taking the raw quantiles of the distribution or by more fancy things like modelling with a skew-normal distribution, because as you can see, the errors are positively skewed and not normally distributed. IF we then apply this empirical forecast error distribution to the point forecasts of our actual forecast, we get empirical prediction intervals.

**14 EMP PIs for Scenarios**

For extrapolative data-driven forecast models, this is very intuitive. But, as you might remember, we also chose a scenario-based approach to forecast Nordic fertility. So, how do we derive empirical prediction intervals for models, that only work under specific circumstances?

**15**

The assumption of the PPS model is that the fertility postponement that has happened so far, will continue but will slow down and stop. In hindsight, it is easy for us to detect other periods where the same has happened and apply our forecast model to them. The fertility postponement already began in the 1970s and there was consensus among researchers that this will slow down and some point. Therefore, we can derive empirical prediction intervals for scenario-based models via cross-validation.

However, there is a limitation. In hindsight, it is easy to apply a model to a time period, where it is certain that it will work. The resulting errors will be small. But this is not reflecting the reality of forecasting. We don’t know, whether our assumption is right and that should be reflected in the prediction intervals!

**16 Publications**

That is why we will also use another approach to retrieve empirical prediction intervals. If we use, like Keilman and Pham also did, published historical forecasts and see how wrong they are we get a different empirical error distribution. One that does reflect the real uncertainty of forecasting, because the published fertility forecasts are the expert consensus of their time.

**17 Results Table**

This was a lot of information and I will try to put this into a more structured way. This is a table of results that we are going to provide. … explain table …

**18**

And this is what we already and what we are currently working on. … explain table …

**19 Preliminary Results**

We now come to the final section, you probably have been waiting for: some results.

**20 TFR PPS**

**21 Finland**

**22 Denmark**

**23 Sweden**

**24 End**

Thank you very much for your attention! I am now happy to discuss this with all of you, and receive your questions, comments and requests.

**25 Norway**