# TO WEIGHT OR NOT TO WEIGHT

## EVALUATION OF THE EFFECT OF WEIGHTING IN THE SWEDISH NATIONAL SOM SURVEY 1991–2022

MS CORNELIA ANDERSSON (PRESENTING)
MS FREJA WESSMAN
DR SEBASTIAN LUNDMARK

# Introduction

- The SOM institute: A university-based research organization, established 1986
  - Conducts yearly national surveys using a simple random sample
  - Broad range of research topics
  - Produces datasets available to researchers and students


- Weights are almost never applied

# Introduction

- Response rates have declined since 1986 in general
  - Response rates are declining at different rates across different population groups

# Introduction

- A lot has happened in the demographic and political landscapes in Sweden.
  - Higher rate within the population of people born outside of Sweden.
  - A new right-winged party has entered the parliament of Sweden.

- This causes concerns that changes in our 40-year long time series are not necessarily reflecting true changes within the population but might instead be caused by some degree of non-response bias.

# Can we trust the changes in our time series?

**Share of people interested in politics, 1991-2022**
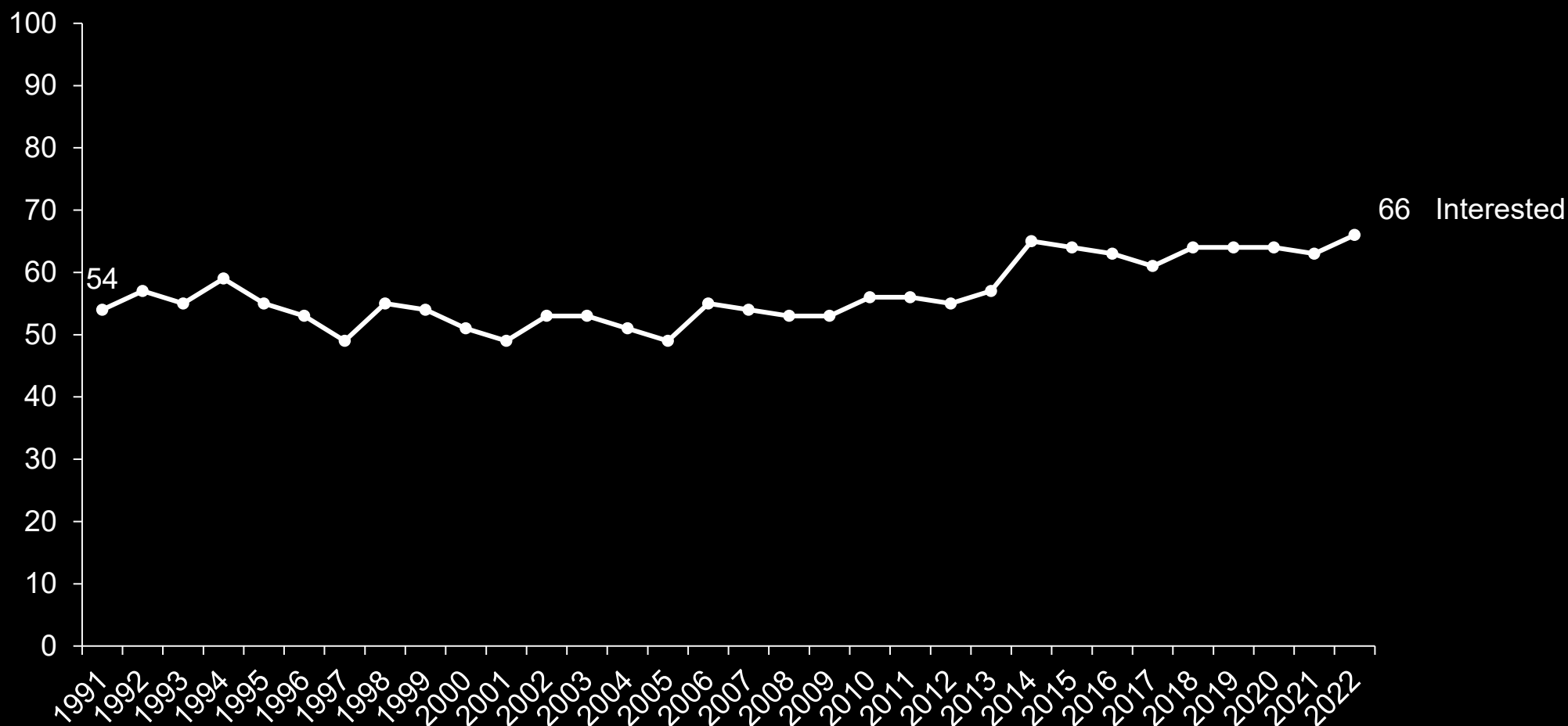
# Purpose of the Project

- Purpose: Determine whether datasets from the national SOM survey should include a general-purpose cell-weight, and, if so, to decide how such a weight should be constructed.

- Phase 1:
  - Identify an optimal cell weight construction that could serve as a general-purpose weight for the annual SOM Survey
    - Optimality is defined as the weight which results in the largest improvement in point estimates compared to true population values

- Phase 2:
  - Assess the performance of this general-purpose weight
    - Apply the optimal cell weight to the time-series data and determine whether routine weighting is warranted
    - If needed, include the weight in the available SOM dataset for researchers and students

# Phase 1 – Evaluate alternative weights' effects on true population estimates

- Population data on…

    - Gender
    - Educational level
    - Metro (resident of a major Swedish city, y/n)
    - Marriage status

    - Age
    - Household income
    - Birth region (Born inside/outside Sweden)

    …was ordered from Statistics Sweden

- A number of initial combinations of cell weights were constructed for each election year from 1991 to 2022

# Phase 1 – Evaluate alternative weights' effects on true population estimates

- True population values for voter turnout, parliamentary election results, and driver's license possession over time were obtained
  - These have been measured in the SOM surveys over time

- Evaluated which weight(s) yielded the greatest improvement on the three estimates over time
  - The evaluation was based on RMSE (root mean square error), a measure of the dispersion compared to a given value

- $\Delta_{RMSE}$ was calculated for each estimate and for each possible year to enable comparison of the individual weight's effects on point estimates.
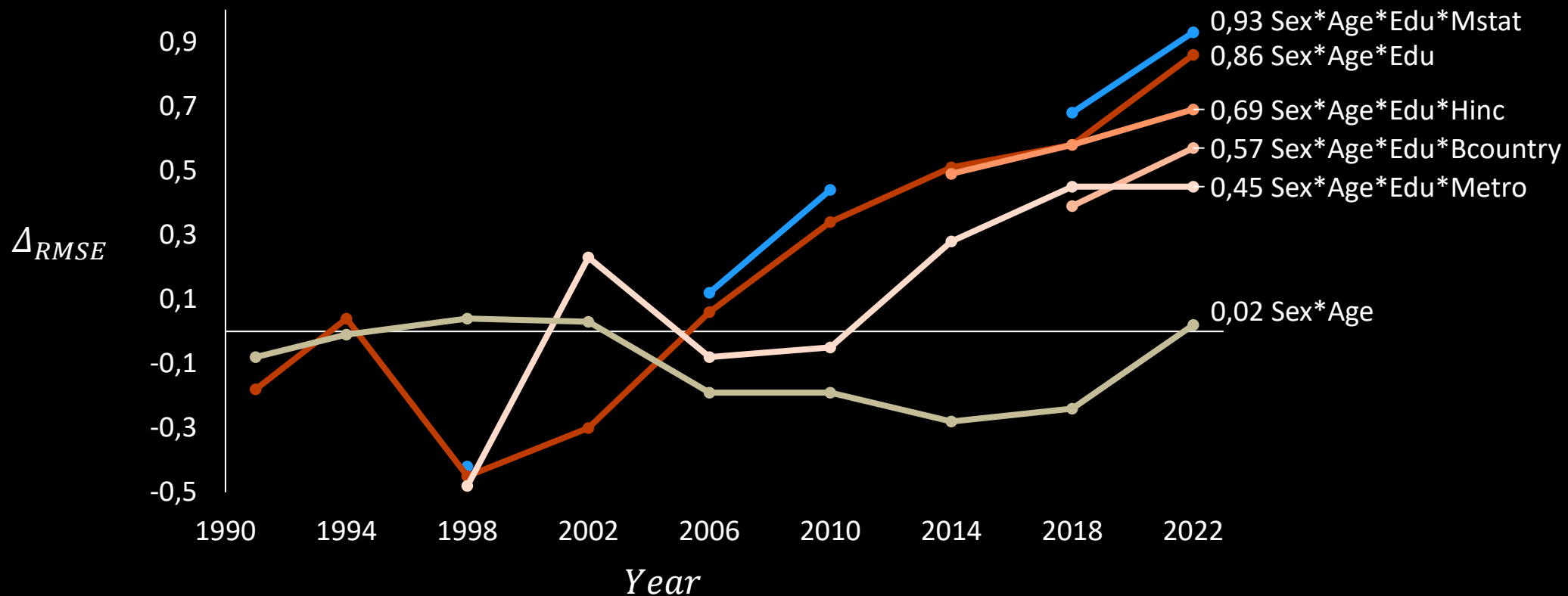  - $\Delta_{RMSE} = RMSE_{weighted} - RMSE_{unweighted}$

Phase 1 – Evaluate alternative weights' effects on true population estimates
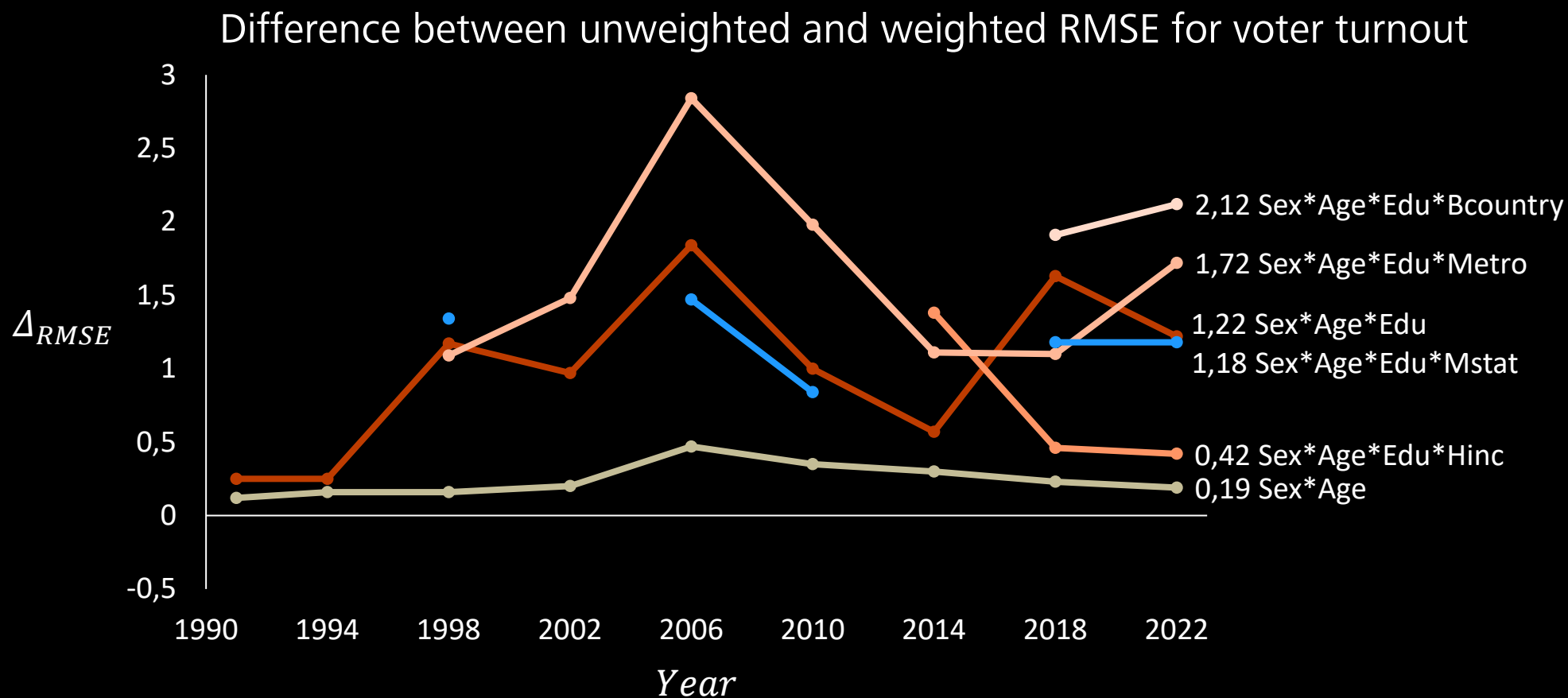
# Phase 1 – Evaluate alternative weights' effects on true population estimates

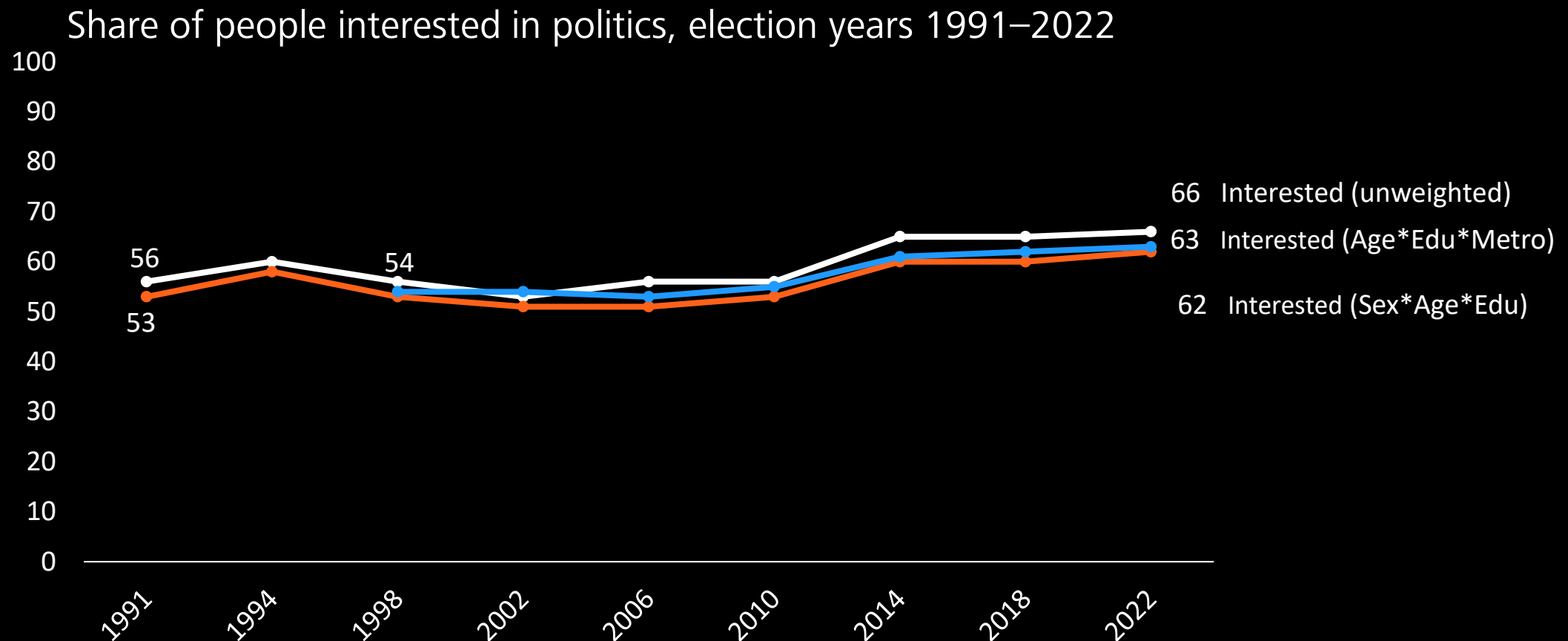Difference between unweighted and weighted RMSE for election results

# Phase 1 – Conclusion

- Most optimal weights overall:
  - **Sex*Age*Edu**
  - **Age*Edu*Metro**

- These weights brought to Phase 2 for further testing

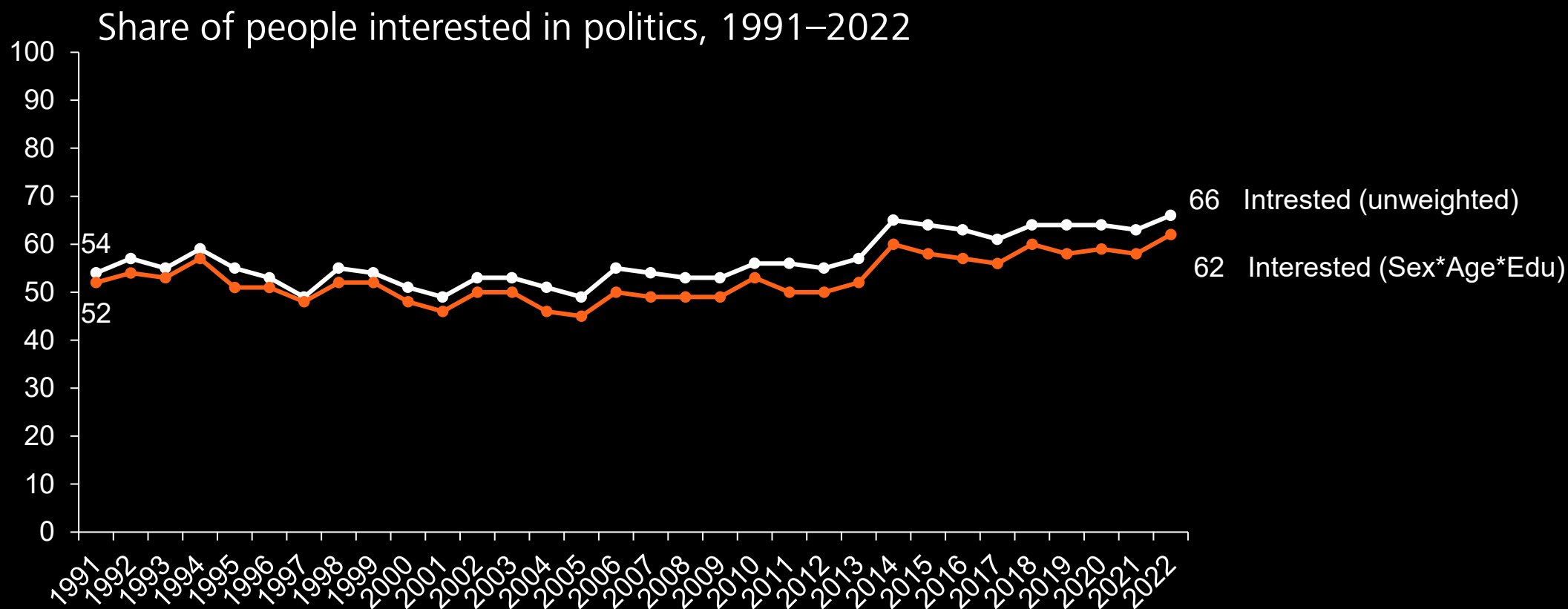# Phase 2 – Applying the two weights on time series from SOM-data

- To assess the impact of the weights in the SOM data, they were applied to ten different variables

Share of people interested in politics, election years 1991–2022



66  Interested (unweighted)

63  Interested (Age*Edu*Metro)

62  Interested (Sex*Age*Edu)

56

54

53

# Phase 2 – Applying one weight on time series from SOM-data

- The weight **sex*age*edu** displayed the overall largest effect on point estimates, and was created for each year 1991 to 2022

Share of people interested in politics, 1991–2022



66   Intrested (unweighted)

62   Interested (Sex*Age*Edu)

# Results

- **Sex*Age*Edu** and **Age*Edu*Metro** had the best overall improvement on measurements when comparing with true population estimates.

- **Sex*Age*Edu** displayed a larger effect when applied on time series, thus potentially correcting a greater portion of non-response bias amongst the respondents.

- However, the effects of weighting were small.
  - Differences between weighted and non-weighted point estimates were often only a few percentage units.

# Conclusion

- No analyzed time series changed direction/pattern when applying the weight(s).

- The small effect of weighting on point estimates provides a risk that the increased variance of the estimate might not make weighting beneficial.

- Self reported measurements from the SOM institute will not be weighted on a general basis.

# Thank you for listening

**Cornelia Andersson**
Deputy Chief Analyst
The SOM institute
University of Gothenburg
cornelia.andersson@som.gu.se