

WP4 - MAKSWELL

Nikos Tzavidis ¹

MAKSWELL Kick-off meeting
Rome, January 9, 2018

¹Southampton Statistical Sciences Research Institute, University of Southampton (n.tzavidis@soton.ac.uk)

WP 4: Overview

Three tasks covering:

- ▶ Task 1: Time series methods for survey estimation with auxiliary information from big data
(CBS-UM/ISTAT/UNIFI)
- ▶ Task 2: Mapping of indicators
(TRIER/CBS-UM/ISTAT/DESTATIS/HCSO/UNIFI)
- ▶ Task 3: Estimating and adjusting for survey discontinuities
(SOTON/TRIER/CBS-UM)

Task 4.1- Combining auxiliary series with series from repeated surveys - Input from CBS/UM

- ▶ Big data sources provide auxiliary series related to series obtained from repeated surveys on a higher frequency
- ▶ Multivariate structural time series (STS) models can be used to combine series from sample surveys with auxiliary series to improve precision of survey sample estimates
- ▶ This is a form of small area estimation to borrow strength over time and from related series
- ▶ The higher frequency of the auxiliary series can also be used to estimate sample survey parameters in real time (nowcasting)
- ▶ **Problems:**
 - ▶ Multivariate STS models assume that correlation between series is time invariant, which is a strong assumptions that is often violated in practise
 - ▶ The number of auxiliary series can be large resulting in high-dimensionality problems

Task 4.1- Combining auxiliary series with series from repeated surveys

- ▶ Statistics Netherlands uses a multivariate STS model to produce official monthly figures about the labour force
- ▶ **Aims:**
- ▶ Extend the model with auxiliary series obtained from claimant counts in search activities on internet
- ▶ Propose methods to allow for time depending correlations (GAS models, nonparametric methods)
- ▶ Propose methods to handle the high-dimensionality problem (e.g. a dynamic factor modelling approach)
- ▶ Mixed frequency approach since LFS is on a monthly frequency and Google trends on a higher frequency
- ▶ Simulation to test aforementioned approaches
- ▶ Application the Dutch LFS

Task 4.2- Mapping of indicators Input from TRIER & DESTATIS

- ▶ Representation of the well-being and performance indicators with maps
- ▶ Mapping change
- ▶ Spatial cluster analysis of extreme indicator values
- ▶ Challenges with interpreting maps
- ▶ **Data:** Micro level survey data on the SDGs, MIP or other performance indicators
- ▶ Information on regional membership

Task 4.3 - Estimating survey discontinuities - Input

from SOTON, CBS/UM and TRIER

- ▶ Surveys try to maintain consistent methodologies (sampling/survey design) over time
- ▶ Aids the comparability of survey estimates over time
- ▶ However, changes in design cannot be avoided
- ▶ Changes designed to increase efficiency/reduce costs
- ▶ Can create breaks in the series known as discontinuities
- ▶ Net effect of all different sources of non-sampling errors changes
- ▶ **Risk:** Producing estimates of interest, e.g. poverty rates that ignore discontinuities
- ▶ To avoid confounding, real period-to-period change with discontinuities, it is important to quantify them
- ▶ Different methods are available e.g. parallel runs and time series modelling

An Example: The National Survey for Wales

- ▶ The Welsh Government (WG) has reviewed the way in which social surveys are conducted in Wales
- ▶ WG instituted a new National Survey (NSn) from 2016
- ▶ The NSn collects information previously collected in 5 surveys
 - ▶ The old National Survey (NSo)
 - ▶ The Welsh Health Survey (WHS)
 - ▶ The Active Adults Survey (AAS)
 - ▶ The Arts in Wales Survey (AWS)
 - ▶ The Welsh Outdoor Recreation Survey (WORS)

Potential Sources of Discontinuities

- ▶ Several changes in the NSn are potentially important
- ▶ Change of contractor
- ▶ Mode: Telephone/self-completed to face-to-face
- ▶ Interviewer effects - Social acceptability (sensitive questions)
- ▶ Questions from 5 surveys combined in a single questionnaire
- ▶ Possible ordering and context effects
- ▶ Impact of new design on response propensity by subgroup

A Framework for Assessing Discontinuities

- ▶ WG put in place a large-scale pilot of the new design
- ▶ Similar design to the one used in the NSn ($n = 2800$)
- ▶ Discontinuities: Difference between the estimates from the old surveys and those from the pilot
- ▶ Take into account the sampling variance
- ▶ Focus on discontinuities greater than 5 percentage points
(Government Statistical Service - Methodology Advisory Committee, 2016)
- ▶ Estimate discontinuities at national and domain (small area) levels

Assumptions

- ▶ **Assumption 1:** The time difference between the pilot and the old surveys can be ignored
- ▶ **Assumption 2:** The pilot is used as if it were the new survey
- ▶ Ideally estimate discontinuities by a split-sample experiment
- ▶ Old and new designs randomly administered to respondents
- ▶ This is not the case with the Welsh pilot survey
- ▶ We cannot say why discontinuities occur

Aims

- ▶ Investigate whether discontinuities present a challenge for other NSIs in Europe
- ▶ Investigate the types of data needed for estimating discontinuities
- ▶ Estimating discontinuities - time series & SAE methods
- ▶ Methods for adjusting series of estimates
- ▶ Applications - Data from the UK and Netherlands
- ▶ Identify areas for methodological research

Deliverables

- ▶ D4.1: Report on time series - M24
- ▶ D4.2: Report on mapping - M29
- ▶ D4.3: Report on survey discontinuities - M24