Timely official statistics during the COVID-19 Pandemic in the Netherlands

Jan van den Brakel[†‡]

†: Statistics Netherlands

‡: Maastricht University

UNECE Expert Meeting on Statistical Data Collection 27-09-2021

Joined work with Sabine Krieg[†] and Marc Smeets[†]

Introduction

Data collection household surveys:

- Sequential mixed-mode design, based on web interviewing (WI), face to face interviewing (CAPI)

COVID-19:

- Partial loss of CAPI during the lockdown in Q2: bias
- Dutch Health Survey (DHS)
  - Annual figures on health related themes
  - Not timely (information published in Q1 2021 will be outdated)
  - No accurate information of the impact of COVID-19 on health figures
- Labour Force Survey
  - Times series model for monthly figures
  - COVID crisis: strong effects on real period-to-period change LFS figures
  - How to separate changes in mode effects from real period-to-period changes?

DHS: COVID-19

- Problem:
  - Loss of CAPI:
    - Sudden change in mode effects (mixture of selection effects and measurement bias)
    - Mode effects will be confounded with the real effect of COVID on health figures
  - Strong external demand on more timely health figures
- Proposed solution:
  - Quarterly figures
  - Requires a model-based inference method
  - Structural time series model:
    - Small area estimation (borrow strength over time)
    - Model is also used to compensate for changing mode effects due to the loss of CAPI

DHS: Structural time series model

Bivariate model:

$$\begin{pmatrix} \hat{y}_t^{\mathrm{C}} \\ \hat{y}_t^{\mathrm{I}} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} (L_t + S_t) + \begin{pmatrix} 0 \\ \lambda_t \end{pmatrix} + \begin{pmatrix} e_t^{\mathrm{C}} \\ e_t^{\mathrm{I}} \end{pmatrix}$$

- Input:
  $\hat{y}_t^{\mathrm{C}}$: Series of direct estimates based on complete response
  $\hat{y}_t^{\mathrm{I}}$: Series of direct estimates based on Internet response only
- STM components
  L_t: smooth trend model
  S_t: trigonometric seasonal model
  λ_t: Random walk for the systematic difference between $\hat{y}_t^{\mathrm{C}}$ and $\hat{y}_t^{\mathrm{I}}$
  e_t^{C} and e_t^{I}: measurement errors of the input series (heteroscedastic and correlated!)
- During the lockdown $\hat{y}_t^{\mathrm{C}}$ is missing
- STM produces nowcasts based on the relation between $\hat{y}_t^{\mathrm{C}}$ and $\hat{y}_t^{\mathrm{I}}$ observed in the past

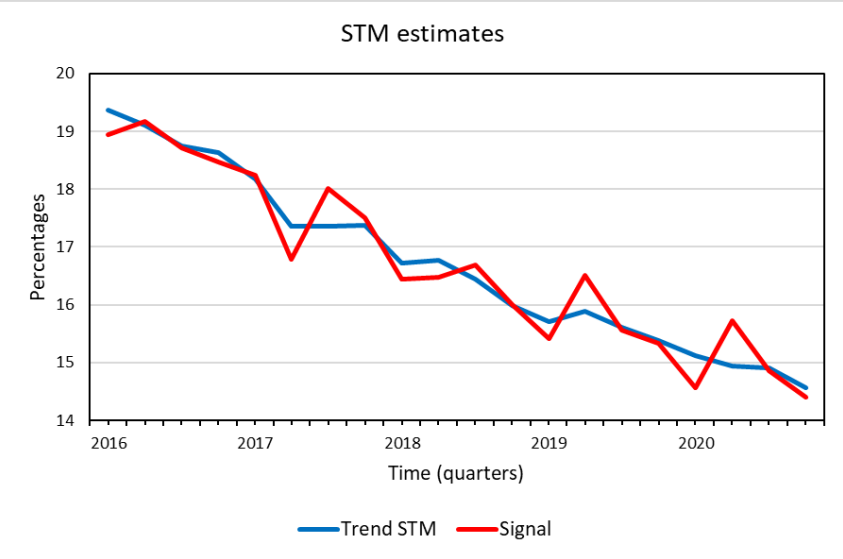
DHS: Results

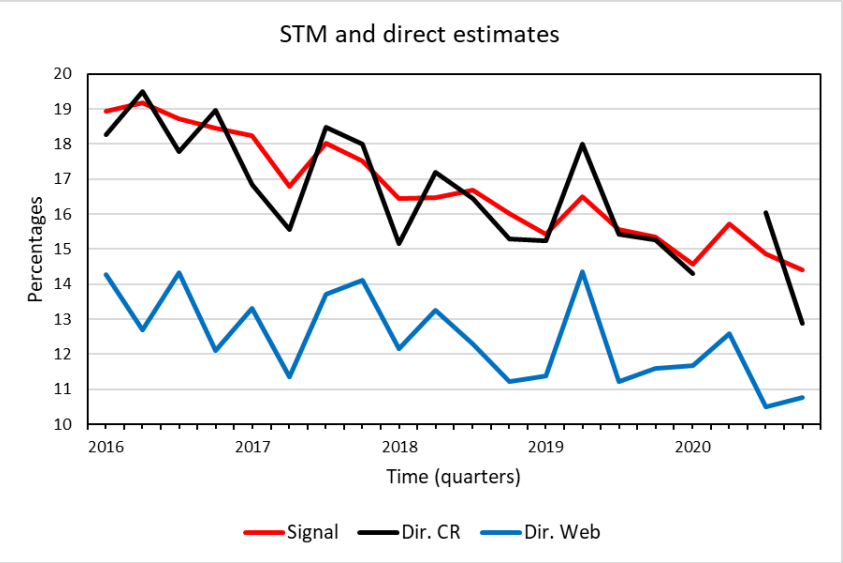
STM's are developed for 8 major indicators:

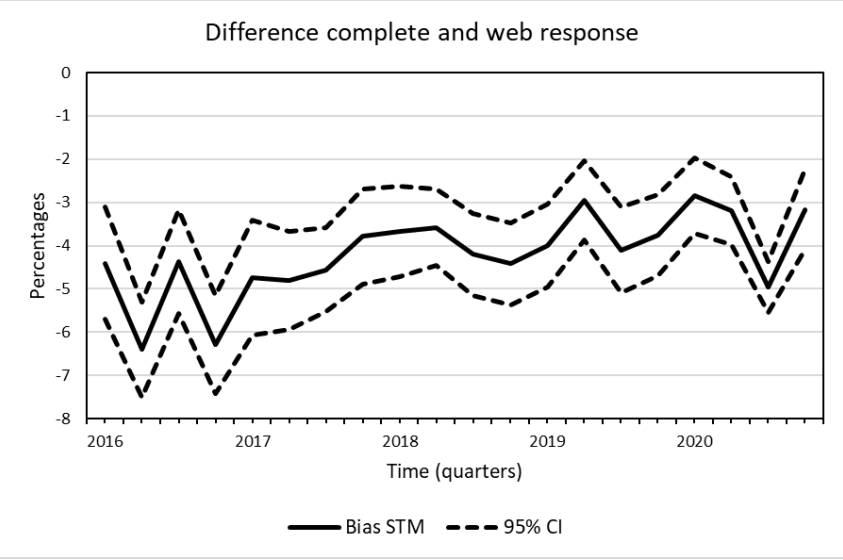
1. Perceived health (0 and older)
2. Mental unhealthy (12 and older)
3. GP visit in the past four weeks (0 and older)
4. Daily smoking (18 and older)
5. Overweight (18 and older)
6. Excessive alcohol consumption (18 and older)
7. Dental visit in the past four weeks (0 and older)
8. Specialist visit in the past four weeks (0 and older)

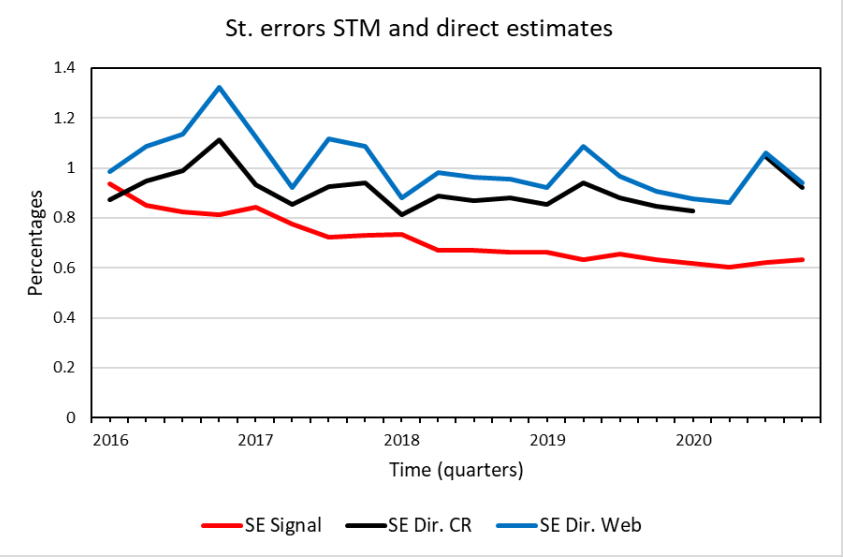
All indicators are defined as percentages

DHS: Results ~ Daily smoking (18 and older)

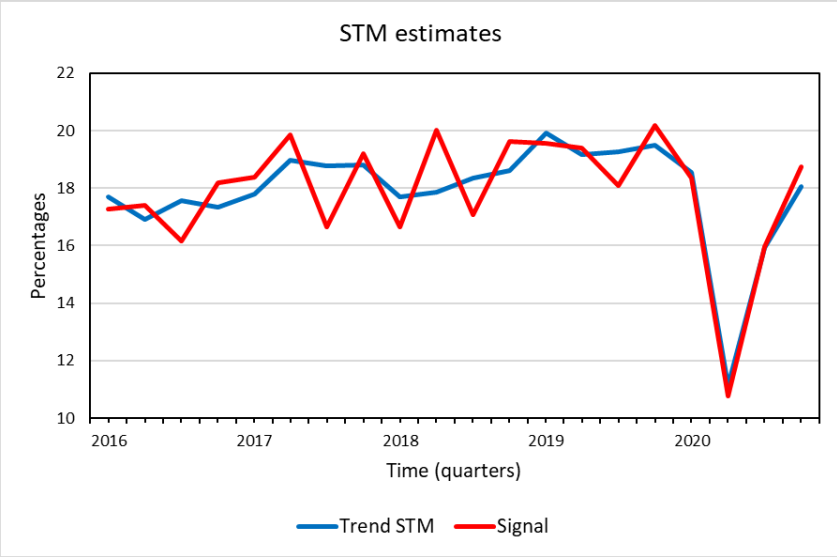


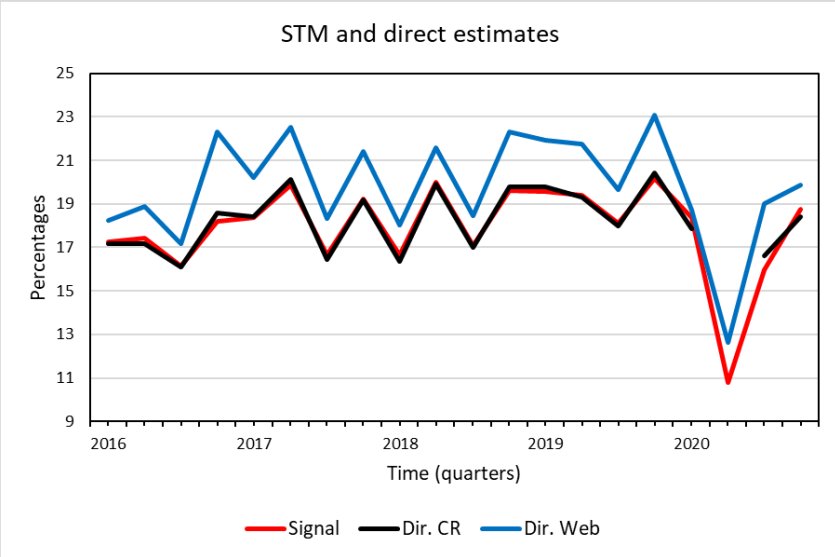


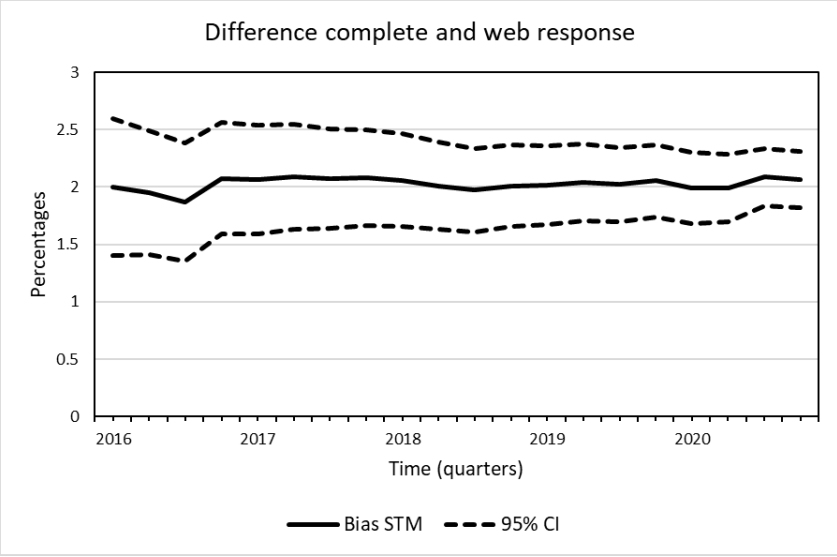


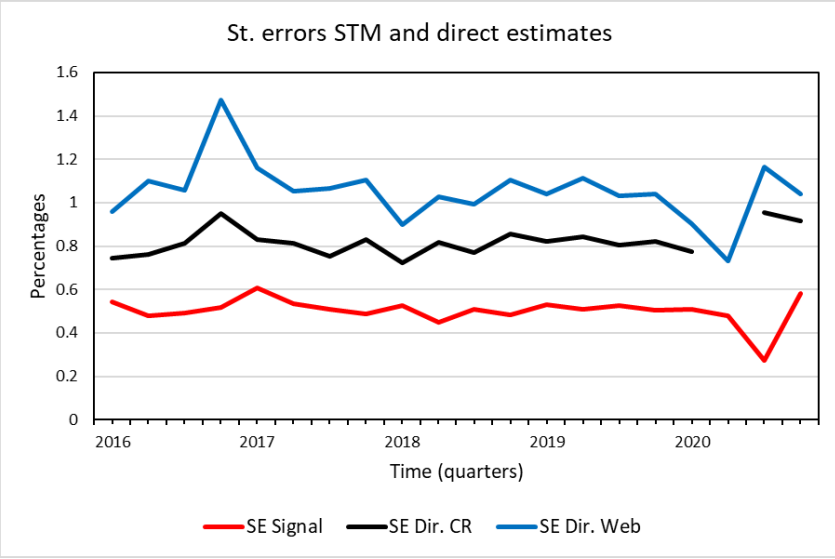


DHS: Results ~ Dental visit in the past four weeks









DHS: Results ~ annual figures

Annual figures

- Based on GREG estimator
- Weighting model is extended with quarterly STM estimates for the 8 variables
  - Numerical consistency between annual and quarterly publications
  - Correction for the loss of CAPI for more detailed breakdowns of the 8 variables
  - A good as possible correction for the loss of CAPI for other related variables for which no STMs are developed

LFS: Estimation procedure for official monthly LFS figures

- Rotating panel design: monthly samples observed 5 times at quarterly intervals
- Structural time series model (implemented in 2010):
  - Small area estimation (borrowing strength from the past)
  - Account for rotation group bias
- Rotation scheme: data collected in 5 independent samples
- $\hat{y}_t^{(j)}$: direct estimate month t, based on the panel that is observed for the j-th time

$$\begin{pmatrix} \hat{y}_t^{(1)} \\ \hat{y}_t^{(2)} \\ \hat{y}_t^{(3)} \\ \hat{y}_t^{(4)} \\ \hat{y}_t^{(5)} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \theta_t^y + \begin{pmatrix} 0 \\ \lambda_t^{(2)} \\ \lambda_t^{(3)} \\ \lambda_t^{(4)} \\ \lambda_t^{(5)} \end{pmatrix} + \begin{pmatrix} e_t^{(1)} \\ e_t^{(2)} \\ e_t^{(3)} \\ e_t^{(4)} \\ e_t^{(5)} \end{pmatrix}$$

population parameter: $\theta_t = L_t + S_t + I_t$ (= trend/cycle + seasonal + white noise).

LFS Problems due to COVID-19

1. No CAPI data collection in the first wave (April until August 2020 and January until June 2021)
   - Shock in $\hat{y}_t^{(1)}$ due to sudden change in mode effects
2. Strong effect on the real period-to-period changes
   - Lockdown marks a sharp turning point in the monthly LFS figures
   - How to account in the time series model for the increased dynamics in the LFS figures

LFS Problem 1: changing mode effects due to loss of CAPI

$$\begin{pmatrix} \hat{y}_t^{(1)} \\ \hat{y}_t^{(2)} \\ \hat{y}_t^{(3)} \\ \hat{y}_t^{(4)} \\ \hat{y}_t^{(5)} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \theta_t^y + \begin{pmatrix} 0 \\ \lambda_t^{(2)} \\ \lambda_t^{(3)} \\ \lambda_t^{(4)} \\ \lambda_t^{(5)} \end{pmatrix} + \begin{pmatrix} \delta_t^{(1)} \beta^{COV} \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} e_t^{(1)} \\ e_t^{(2)} \\ e_t^{(3)} \\ e_t^{(4)} \\ e_t^{(5)} \end{pmatrix}$$

with $\theta_t \;\; = L_t + S_t + I_t$

$\hat{y}_t^{(1)}$: based on WI and CATI starting from April until August 2020 and from January until June 2021

$\delta_t^{(1)} = \begin{cases} 1 & \text{from April until August 2020 \& Januari until June 2021} \\ 0 & \text{otherwise} \end{cases}$.

β^{COV}: discontinuity for the difference between $\hat{y}_t^{(1)}$ with and without CAPI

estimated with a separate time series model: next slide

LFS Problem 1: changing mode effects due to loss of CAPI

Estimation β^{COV}: difference first wave with and without CAPI

- Used series first wave observed in the past under the sequential mixed mode design implemented in April 2012
- $\hat{y}_t^{(1)}$: series direct estimates first wave based on WI, CATI and CAPI from April 2012 – March 2020
- $\hat{y}_t^{(*1)}$: series direct estimates first wave based on WI and CATI only from April 2012 – March 2020
- x_t : series of claimant counts

$$\begin{pmatrix} \hat{y}_t^{(1)} \\ \hat{y}_t^{(*1)} \\ x_t \end{pmatrix} = \begin{pmatrix} L_t^y \\ L_t^y \\ L_t^x \end{pmatrix} + \begin{pmatrix} 0 \\ \Delta_t^{Ly} \\ 0 \end{pmatrix} + \begin{pmatrix} S_t^y \\ S_t^y \\ S_t^x \end{pmatrix} + \begin{pmatrix} e_t^{y1} \\ e_t^{y1*} \\ e_t^x \end{pmatrix}$$

$$\beta^{COV} \equiv \Delta_T^{Ly}$$

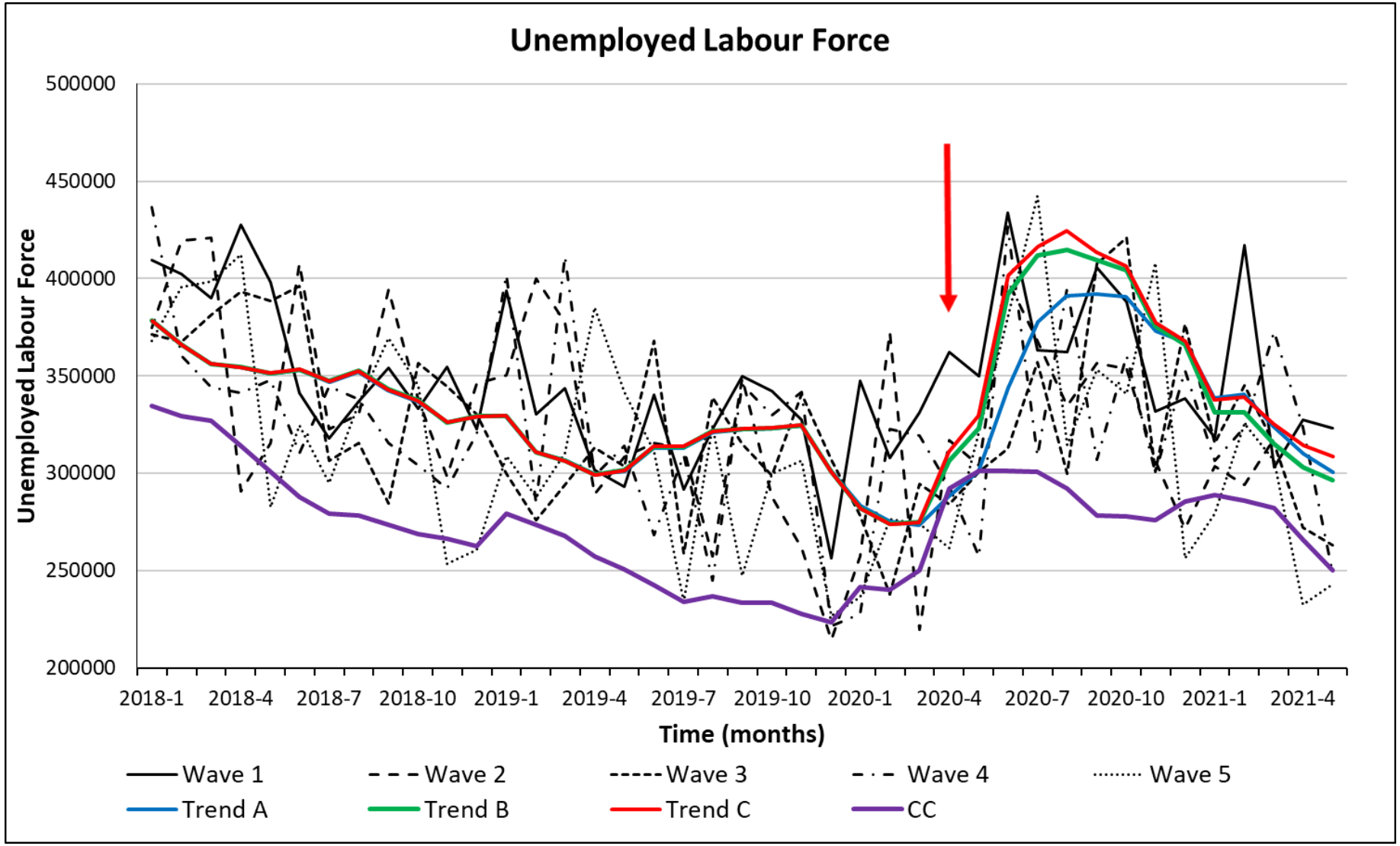
LFS Problem 2: increased dynamics in the LFS figures

Time series component for the population parameter:

$$\theta_t^y = L_t + S_t + I_t$$

Trend:

- Not flexible enough to follow the sudden increase in period-to-period changes in April 2020
- Modify the model to increase the flexibility of the trend
- Interpretation:
  - By increasing the model variance, the time series model gives more weight to the direct estimates in April and less weight to the model predictions based on the past

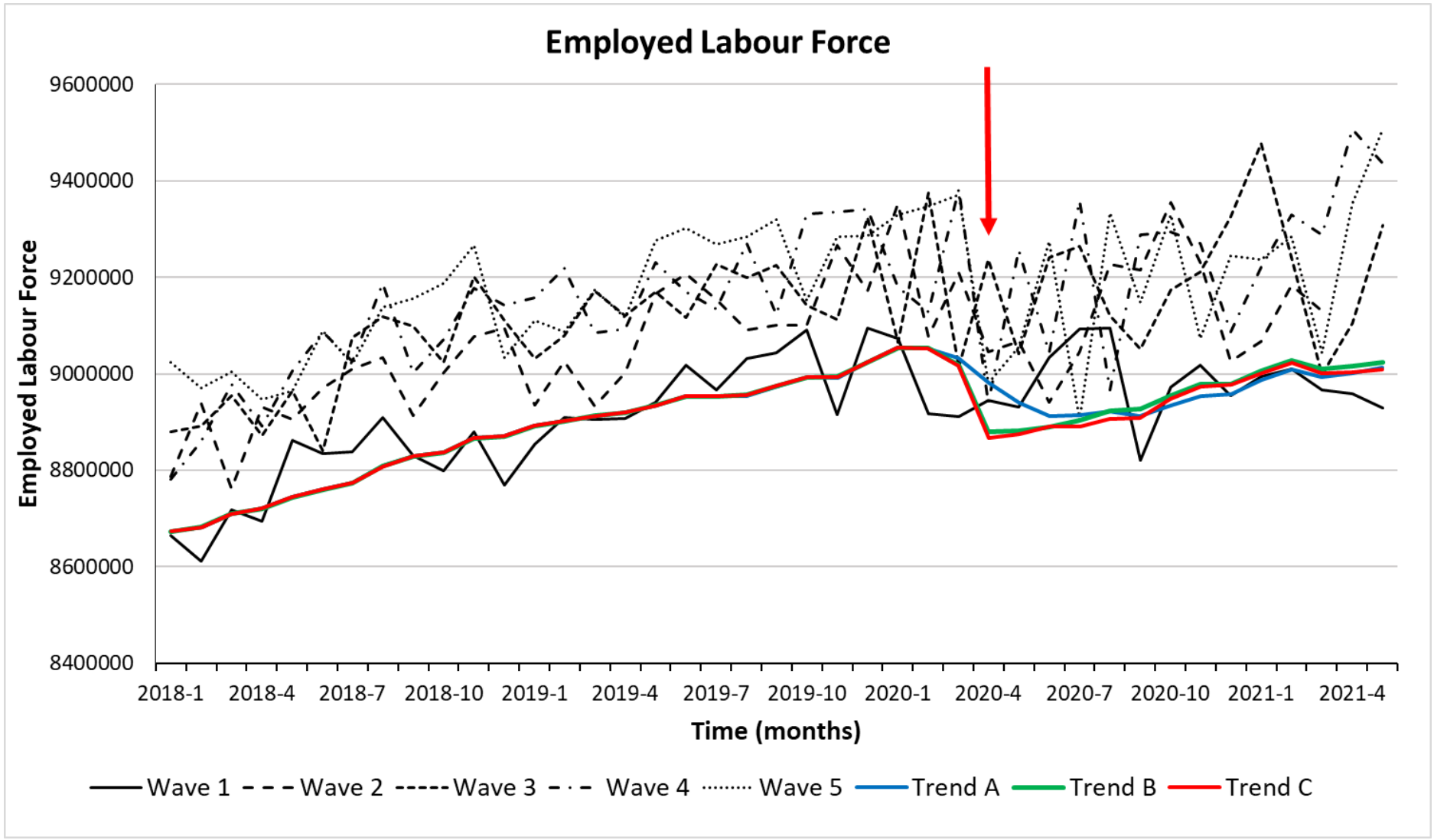


Trend A: unadjusted production model

Trend B: time varying trend variance

Trend C: time varying trend variance + correction missing CAPI

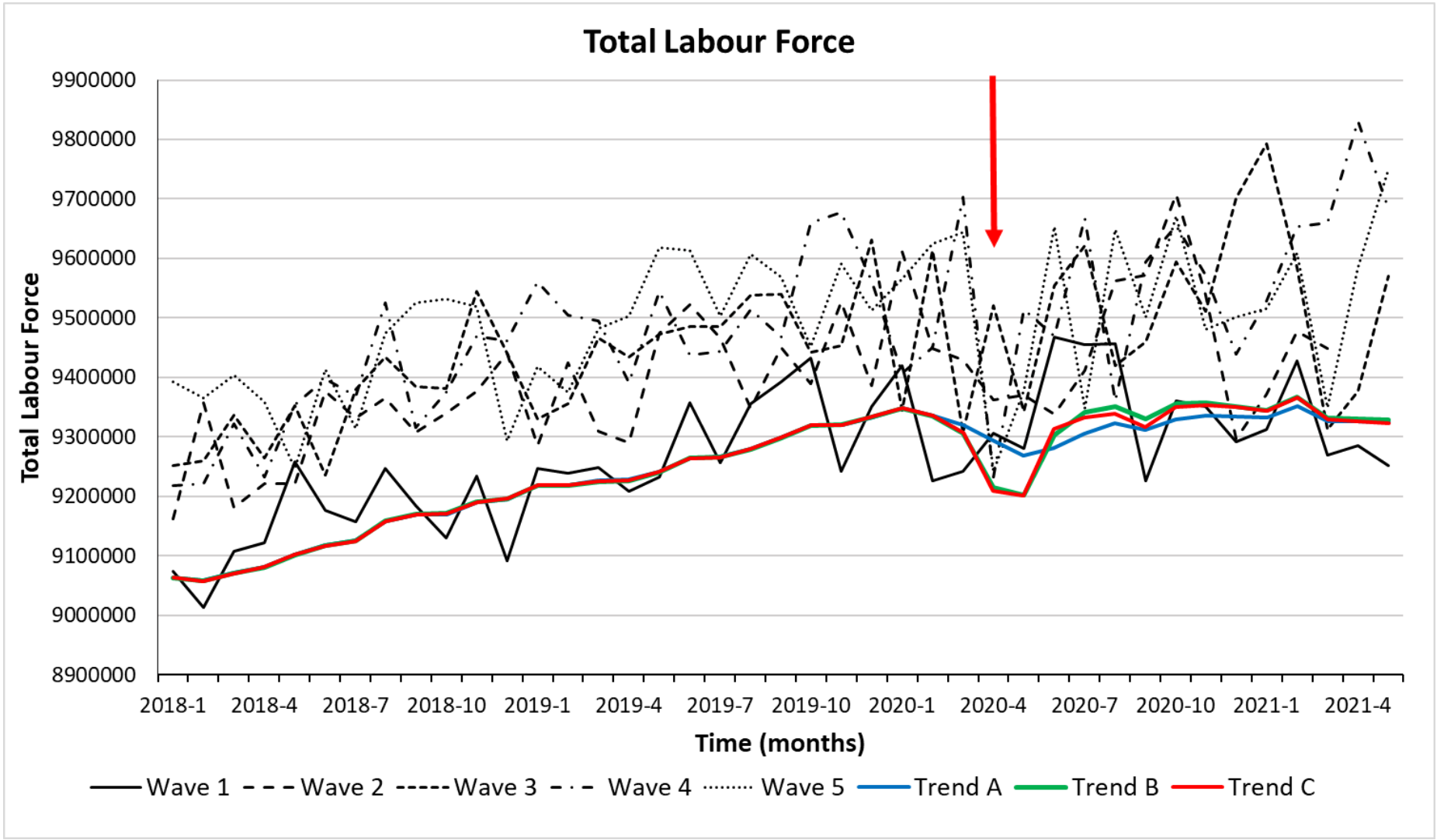
CC: claimant counts



Trend 1: unadjusted production model

Trend 2: time varying trend variance

Trend 3: time varying trend variance+correction missing CAPI



Trend 1: unadjusted production model

Trend 2: time varying trend variance

Trend 3: time varying trend variance+correction missing CAPI

Discussion

- Multivariate STM to compensate for loss of CAPI and to produce more timely figures
- Bias correction is based on a strong assumption that cannot be verified:
  - Observed difference between $\hat{y}_t^{\mathrm{C}}$ and $\hat{y}_t^{\mathrm{I}}$ is not affected by the lockdown
  - Internet respondents before and during the lockdown are comparable

Details:

- J.A. van den Brakel, M. Souren and S. Krieg (2021). Estimating monthly Labour Force Figures during the COVID-19 pandemic in the Netherlands. Discussion paper, February, 2021, Statistics Netherlands, Heerlen.

- Email: jbrl@cbs.nl