Estimating monthly Labour Force Figures during the COVID-19 pandemic in the Netherlands

Keywords: Structural time series models, discontinuities, turning points, monthly labour force figures

1. INTRODUCTION

Official monthly statistics about the Dutch labour force are based on the Dutch Labour Force Survey (LFS). The LFS is a continuously conducted survey that is designed as a rotating panel. Data collection among selected households is based on a mixed-mode design based on web interviewing (WI), telephone interviewing (CATI) an face-to-face interviewing (CAPI). Monthly estimates about the labour force are obtained with a structural time series model.

Due to the COVID-19 pandemic, the Netherlands went in a lockdown on March 16, 2020. Due to this lockdown, face-to-face interviewing stopped. It is anticipated that this has a systematic effect on the outcomes of the LFS. At the same time, it can be expected that the lockdown effects the real monthly labour force figures. The lockdown indeed marked a sharp turning point in the evolution of these series and strongly increases the dynamics in the figures of the labour force series.

In this paper it is explained how Statistics Netherlands produces monthly labour force figures during the COVID-19 pandemic. It is shown how the sudden change in the mode effects, because face-to-face interviewing stopped, is separated from real period-to-period changes in the labour force figures. It is also explained how the time series model is adapted to the increased dynamics in the labour force figures.

2. METHODS

2.1. Time series for the LFS

The Dutch LFS is based on a stratified two-stage cluster sample of addresses. Each month a new sample enters the panel, which is observed five times at quarterly intervals. The sample leaves the panel after the fifth interview. Data collection in the first wave is based on a mixed-mode design using WI, CATI and CAPI. Data collection in the follow-up waves is based on CATI only. Since 2010, monthly figures about the labour force are produced with a structural time series model. This model-based estimation procedure solves two problems. First it used as a form of small area estimation to account for the relative small monthly sample sizes, which hampers the use of direct estimators. Second it accounts for systematic differences between the outcomes of the subsequent waves of the panel, in the literature known as rotation group bias (RGB).

According to the rotation scheme of the panel design, households are interviewed five times at quarterly intervals. This implies that each month data are collected in five independent samples. Let $\hat{y}_t^j$ denote the general regression (GREG) estimate for the unknown population parameter, say θ_t, based on the *j*-th panel observed at time *t*, (*j* = 1, …, 5). The GREG estimates of each month can be expressed as a vector $\widehat{\boldsymbol{y}}_t = (\hat{y}_t^1, \dots, \hat{y}_t^5)^t$, which are the input series for the following five dimensional structural time series model:

$$\widehat{\boldsymbol{y}}_t = \boldsymbol{j}_5(L_t + S_t + I_t) + \boldsymbol{\lambda}_t + \boldsymbol{e}_t \qquad (1)$$

The first component in (1) is the Basic Structural Time Series Model for the unknown population parameter. It decomposes the unobserved true value of the population parameter θ_t in a dynamic trend (L_t), a seasonal component (S_t) and a white noise component (I_t) for the unexplained variation. In this first component $\boldsymbol{j}_5$ is five dimensional vector with each element equal to one, since the five GREG estimates in $\widehat{\boldsymbol{y}}_t$ are five independent estimates for θ_t. For L_t the so-called smooth trend model and for S_t a tri-goniometric seasonal model is assumed. See [1] for details.

The second component in (1), i.e. $\boldsymbol{\lambda}_t = (\lambda_t^1, \dots, \lambda_t^5)^t$, accounts for systematic differences between the outcomes of the subsequent waves of the panel, i.e. the RGB. It is understood that superscripts refer to the wave number and should not be interpreted as powers. The absolute bias cannot be estimated from the sample data only. Therefore, additional restrictions for the elements of $\boldsymbol{\lambda}_t$ are required to identify the model. Here it is assumed that an unbiased estimate for θ_t is obtained with the first panel. This implies that the first component of $\boldsymbol{\lambda}_t$ equals zero. The other measure the time dependent differences with respect to the first panel and are modelled as random walks. As a result $\boldsymbol{\lambda}_t = (0, \lambda_t^2, \dots, \lambda_t^5)^t$, with λ_t^j, $j = 2, \dots, 5$, four different random walks.

The third component in (1), $\boldsymbol{e}_t = (e_t^1, \dots, e_t^5)^t$, models the sampling errors of the input series and account for heteroscedasticity due to varying sample sizes over time and serial auto correlation due to the partial sample overlap of the rotating panel design. See [2] for details of this model and its use in the production of monthly labour force figures.

2.2 Impact of COVID-19

The lockdown at 16th March 2020 in the Netherlands had two major effects. First, CAPI data collection stopped. Suddenly stopping CAPI data collection results in a sudden change of measurement bias and selection bias in the responses of the LFS and therefore has a systematic effect on the sample estimates. Second, the lockdown has a strong effect on the real evolution of the labour market figures. After a period of seven years of steady decline of the unemployed labour force and a steady increase of the employed labour force, the lockdown indeed marks a sharp turning point in the time series of both figures.

A discontinuity in the series of the LFS due to suddenly stopping the CAPI data collection could be modelled with a level intervention. Extending model (1) with an intervention component, however, is in this case not a potential solution to account for the discontinuity due to the loss of CAPI respondents because the lockdown also has a strong effect on the real period-to-period change of the labour force figures. In this case it can be expected that a substantial part of the turning point is the series is incorrectly absorbed in the level intervention.

As an alternative, the discontinuity due to the loss of CAPI is estimated with a separate multivariate structural time series model that uses as input:

- Time series of GREG estimates of the first wave based on WI, CATI and CAPI observed up until March 2020
- Time series of GREG estimates of the first wave based on WI and CATI only observed up until March 2020
- Time series of claimant counts observed up until March 2020

Subsequently, the estimate for the discontinuity is used as a priori information in model (1), extended with a level intervention, i.e.:

$$\widehat{\boldsymbol{y}}_t = \boldsymbol{j}_5(L_t + S_t + I_t) + \boldsymbol{\lambda}_t + \Delta_t + \boldsymbol{e}_t \qquad (2)$$

with $\Delta_t = (\delta_t \beta, 0, 0, 0, 0)^t$ where δ_t is a dummy indicator which is equal to one during the period that CAPI response is missing in the first wave and zero elsewhere, and β an estimate for the discontinuity due to the loss of CAPI estimated with the aforementioned separate multivariate structural time series model. The estimate for β is used in the analysis of model (2) via an exact initialization of the Kalman filter.

An additional effect of the lockdown is a strong increase in the dynamics of the input series. In April it became clear that model (2) was miss-specified because it could not capture the sharp turning point in the evolution of the population parameter. This miss-specification was detected with extremely large standardized innovations for the first months under the lockdown. To account for the increased dynamics in the input series in the time series model, the variance of the trend component is made time dependent. As mentioned in Subsection 2.2, the trend is modelled with a smooth trend model, i.e.:

$$L_t = L_{t-1} + R_{t-1}, \qquad R_t = R_{t-1} + \eta_t, \qquad (3)$$
$$\eta_t \cong N(0, \sigma_\eta^2),$$

In (3), R_t is interpreted as the slope parameter of the trend. The flexibility of this trend model is determined by the variance component σ_η^2 of the slope disturbance terms η_t in (3). The flexibility of the trend component is increased during the start of the lockdown by multiplying the variance component with a time dependent factor, i.e.:

$$\eta_t \cong N(0, k_t \sigma_\eta^2), \qquad (4)$$

with $k_t = 1$ for the period before the lockdown and $k_t > 1$ during the first periods of the lockdown. Values for k_t are chosen such that:

- Standardized innovations have reasonable values (between 2 and 2.5 in absolute values)
- The in real time maximum likelihood estimates for σ_η^2 remain stable before and after the start of the lockdown

3. RESULTS

Table 1 presents the discontinuities of the monthly labour force figures estimated with the additional time series model described in Subsection 2.2. Point estimates are made consistent with a Lagrange function, such that the sum of the employed and unemployed labour force equals the total labour force and the sum over the domains equals the national totals.

	Unemployed L.F.		Employed L.F.		Total L.F.	
National	-20,848	(2,555)	35,052	(4,206)	14,203	(4,125)
Men 15-24	-8,494	(3,023)	-9,903	(2,052)	-18,396	(2,128)
Women 15-24	-7,097	(3,469)	-8,140	(1,950)	-15,236	(4,097)
Men 25-44	-11,539	(1,059)	29,000	(1,851)	17,461	(4,097)
Women 25-44	-5,268	(884)	36,243	(2,711)	30,976	(8,651)
Men 45-65	5,739	(1,083)	-3,325	(6,413)	2,415	(6,543)
Women 45-65	5,809	(2,182)	-8,824	(11,774)	-3,015	(10,273)

Table 1: Discontinuity estimates due to the loss of CAPI during the COVID-19 lockdown (standard errors in brackets)

The estimates in Table 1 are used as a priori information to initialize β in model (2). This achieved with an exact initialization of the Kalman filter for this particular state variable.

Figure 1 compares the GREG estimates of the five input waves from January 2018 until September 2020 with:

- The filtered trend under the unadjusted production model (1) (trend 1 in red)
- The filtered trend under model (1), where the variance of the trend component is time dependent using model (4) (trend 2 in green)
- The filtered trend under model (2), where the variance of the trend component is time dependent using model (4) and the model corrects for the discontinuity due to the loss of CAPI (trend 3 in brown)

- Claimant count series available from a register (CC in purple)

Statistics Netherlands official publications are since April 2020 based on model (2) that accounts for a more flexible trend and for the loss of CAPI response (i.e. Trend 3 in brown in Figure 1).

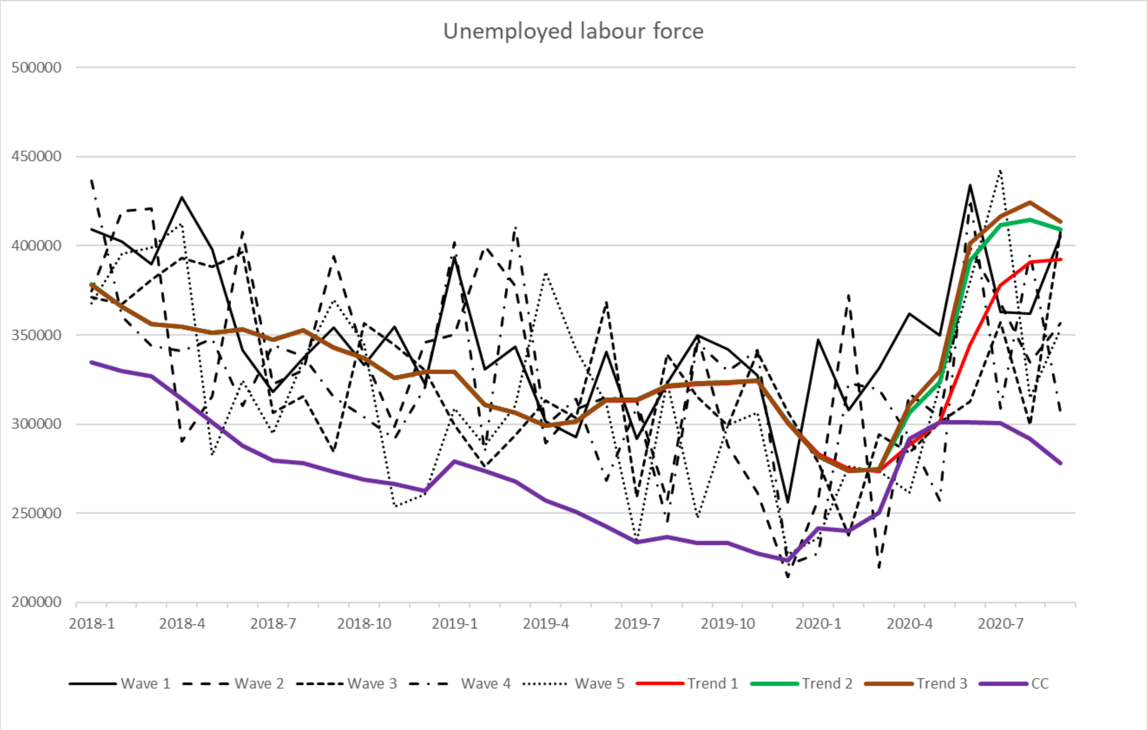

Figure 1: Unemployed labour force at national level, input series, trend estimates under three different models and claimant counts

4. CONCLUSIONS

Statistics Netherlands applies a structural time series model to produce official monthly labour force figures. During the lockdown, this model is extended to account for discontinuities due to the sudden loss of CAPI response in the first wave and the sudden increase of the dynamics in the input series. It can be concluded that the impact of the sudden increase in the dynamics of the input series is much larger compared to the impact due to the loss of CAPI response. Accounting for the loss of CAPI response is, nevertheless, important for the underlying domains. Without adjusting the production model for both effects of the COVID-19 crisis, the model be miss-specified during the lockdown resulting in a severe underestimation of the unemployment figures.

REFERENCES

[1] Durbin, J. and S.J. Koopman (2012). *Time Series Analysis by State Space Methods*. Oxford University Press.

[2] Van den Brakel, J.A. and S. Krieg (2015). Dealing with small sample sizes, rotation group bias and discontinuities in a rotating panel design. *Survey Methodology*. Vol. 41, pp. 267-296.