Discussion Paper

Transition plan for the redesign of the Dutch Labour Force Survey

Jan van den Brakel and Sumonkanti Das

June 19, 2020

Monthly figures about the Dutch Labour Force are based on the Labour Force Survey. Official figures are produced with a multivariate structural time series model, which is used as a form of small area estimation to improve the precision of monthly estimates and to account for rotation group bias and discontinuities induced by two redesigns that toke place in 2010 and 2012. Based on a Eurostat regulation, Statistics Netherlands is preparing a third redesign that is planned for 2021. In a well-designed transition process discontinuities are quantified to avoid confounding real period-to-period change from the systematic effect of the redesign on the outcomes of the survey estimate. One way to quantify discontinuities is conduct a parallel run, where data are collected under the old and new survey designs alongside each other for some period. Because of limited budget and field capacity a large parallel run is not possible for the upcoming transition. As an alternative, the information obtained with a relative small parallel run is combined with a time series modeling approach, where discontinuities are estimated with a level intervention. In this paper simulations under different transition scenarios are conducted to obtain insight with which precision discontinuities as well as the trend under the old and new level during the transition can be estimated.

Key words: Discontinuities, structural time series models, survey redesign

This work was funded by the European Union under grant no. 826475 - 2018-NL-LFS QUALITY BREAKS.

1 Introduction

Official monthly statistics about the Dutch labour force are based on the Dutch LFS. This survey is based on a rotating panel design. The responding households are interviewed five times at quarterly intervals, which implies that every month five waves are being interviewed. The estimation procedure of the LFS is based on the GREG estimator. In 2010, Statistics Netherlands implemented a model-based estimation procedure based on multivariate structural time series model to produce monthly figures about the labour force (Pfeffermann, 1991; van den Brakel and Krieg, 2015). With this time series model, three substantial problems are solved. The first problem is that the monthly sample size of the LFS is too small to rely on the GREG estimator to produce timely official monthly statistics about the employed and unemployed labour force. The time series model uses information observed in previous periods to improve the effective sample size of the individual months to obtain more precise sample estimates as a form of small area estimation. The second problem is that there are substantial systematic differences between the subsequent waves due to mode and panel effects. This is a well-known problem for rotating panel designs, and in the literature this is referred to as rotation group bias (RGB), (Bailar, 1975). At the moment the LFS changed from a cross-sectional survey to a rotating panel design in October 1999, the effects of the RGB on the outcomes of the LFS became very visible. The time series model accounts for RGB by benchmarking the results observed in the follow-up waves to the level of the first wave. The third problem is the systematic effect on the outcomes of the LFS due to two major redesigns of the survey process in 2010 and 2012. Redesigns generally affect the various non-sampling error sources in a survey process, and therefore result in systematic effects on the outcomes of a survey. In an ideal survey transition process, these so-called discontinuities are quantified in order to keep series consistent and preserve comparability of the outcomes over time. In both redesigns, discontinuities are quantified using parallel data collection and modeling discontinuities in the time series model using intervention variables (van den Brakel and Roels, 2010).

Based on a Eurostat regulation, a third redesign of the Dutch LFS will be implemented in 2021. The purpose of this paper is to describe a transition plan to implement the new survey design of the LFS which avoids the risk of disrupting the continuity of time series of labour force figures. In particular, extensive simulation are conducted under different transition scenario's to establish with which precision discontinuities can be estimated as well as the labour force figures under the old and new design. These simulations give better insight how to design the transition.

In Section 2, the survey design of the Dutch Labour force survey is described. The time series model used to produce official monthly figures about the labour force is described in Section 3. The redesign that is planned for 2021 is described in Section 4. The simulation set up is proposed in Section 5 and results are presented in Section 6. The paper concludes with a discussion in Section 7.

2 Dutch Labour Force Survey

The objective of the Dutch LFS is to provide reliable information about the Dutch labour force. Each month a stratified two-stage cluster sample of addresses is drawn. Strata are formed by geographical regions. Municipalities are considered as primary and addresses as secondary sampling units. All households residing at an address, up to a maximum of three, are included in the sample. Different subpopulations are oversampled to improve the accuracy of the official releases, for example, addresses where people live, who are formally registered at the employment office, and subpopulations with low response rates. Before 2000, the LFS was designed as a cross-sectional survey. Since October 1999, the LFS has been conducted as a rotating panel design. Until the redesign in 2010, data in the first wave were collected by means of computer assisted personal interviewing (CAPI). Respondents were re-interviewed four times at quarterly intervals by means of computer assisted telephone interviewing (CATI). During these re-interviews, a condensed questionnaire was used to establish changes in the labour market position of the respondents.

In 2010, a major redesign for the LFS started. The main objective of this redesign was to reduce the administration costs of this survey. This is accomplished by changing the data collection in the first wave from CAPI to a mixed data collection mode using CAPI and CATI. Households with a listed telephone number are interviewed by telephone, the remaining households are interviewed face-to-face. To make CATI data collection in the first wave feasible, the questionnaire for the first wave needed to be abridged since a telephone interview, according to data collection literature, should not take longer than 15 to 20 minutes. Therefore, parts of the questionnaire were transferred from the first to the second or the third wave.

In 2012, a second major redesign of the LFS took place. Data collection changed to a sequential mixed-mode design that starts with Web interviewing. After three reminders the non-responding households are contacted by telephone if they have a listed telephone number. The remaining households are interviewed face-to-face. It was again necessary to change the questionnaire in all the five waves.

The monthly gross sample size for the first wave averaged about 8,000 addresses commencing the moment that the LFS changed to a rotating panel design and gradually fell to about 6,500 addresses in 2012. The response rate is about 55% in the first wave and in the subsequent waves about 90% with respect to the responding households from the preceding wave. After the second redesign in 2012, the monthly sample size in the first wave increased to about 8500 households. Response rates in the first wave vary between 50% and 55%, in the second wave about 70% of the households respond with respect to the responding households from the first wave and in the third, fourth and fifth waves about 90% with respect to the responding households from the preceding wave.

3 Time series model for monthly labour force figures

As mentioned in the introduction, monthly figures about the labour force are based on a time series model. The estimation procedure of the LFS starts with the general regression (GREG) estimator (Särndal et al., 1992). Inclusion probabilities reflect the sampling design and differences in response rates between geographic regions. The weighting scheme is based on a combination of different socio-demographic categorical variables. Key parameters of the LFS are the employed, unemployed and total labour force, which are defined as population totals. Another important parameter is the unemployment rate, which is defined as the ratio of the unemployed labour force to the total labour force.

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 GREG estimate for the unknown population parameter, say θ_t , based on the j-th wave observed at time t, $j = 1, ..., 5$. Since responding households are interviewed at quarterly intervals, it follows that the j-th wave at time t that was sampled for the first time at time $t - 3j + 3$. The GREG estimates of each month can be expressed as a vector $\hat{\mathbf{y}}_t = (\hat{y}_t^{[1]}, \ldots, \hat{y}_t^{[5]})^t$. A five dimensional time series with GREG estimates for the monthly employed and unemployed labour force is obtained as a result. Pfeffermann (1991) proposed a multivariate structural time series model for this kind of time series to model the population parameter of interest, and to account for the RGB and the autocorrelation in the sampling errors. This approach is extended with an intervention component to model the discontinuities of the survey redesign. This results in the following time series model for the five series of GREG estimates:

$$\hat{\mathbf{y}}_t = \mathbf{j}_{[5]}\theta_t + \Lambda_t + \Delta_t^{[1]}\mathrm{B}_t^{[1]} + \Delta_t^{[2]}\mathrm{B}_t^{[2]} + \mathbf{e}_t, \tag{1}$$

with $\mathbf{j}_{[5]}$ a five dimensional vector with each element equal to one, $\Lambda_t = (\lambda_t^{[1]}, \lambda_t^{[2]}, \lambda_t^{[3]}, \lambda_t^{[4]}, \lambda_t^{[5]})^t$ a vector with time dependent components that account for the RGB, $\Delta_t^{[i]} = Diag(\delta_t^{[i,1]}, \delta_t^{[i,2]}, \delta_t^{[i,3]}, \delta_t^{[i,4]}, \delta_t^{[i,5]})$ a diagonal matrix with dummy variables that change from zero to one at the moment that the survey changes from the old to the new design during redesign $i = 1$ in 2010 and $i = 2$ in 2012, $\mathrm{B}_t^{[i]} = (\beta_t^{[i,1]}, \beta^{[i,2]}, \beta^{[i,3]}, \beta^{[i,4]}, \beta^{[i,5]})^t$ a five dimensional vector with regression coefficients, and $\mathbf{e}_t = (e_t^{[1]}, e_t^{[2]}, e_t^{[3]}, e_t^{[4]}, e_t^{[5]})^t$ the corresponding survey errors for each wave estimate. The information obtained during the parallel run in the first wave is used as a-priori information in the model and therefore the regression $\beta_t^{[i,1]}$ is made time varying during the period of the parallel run.

The population parameter θ_t in (1) can be decomposed in a trend component, a seasonal component, and an irregular component, i.e.

$$\theta_t = L_t + S_t + \epsilon_t \tag{2}$$

L_t in (2) is the so-called smooth trend model, which is defined as

$$\begin{aligned} L_t &= L_{t-1} + R_{t-1} \\ R_t &= R_{t-1} + \eta_t \qquad \eta_t \simeq N(0, \sigma_\eta^2) \end{aligned}$$

Furthermore, S_t in (2) denotes a trigonometric stochastic seasonal component, which is defined as

$$\begin{aligned} S_t &= \sum_{j=1}^{J/2} \gamma_{j,t} \\ \gamma_{j,t} &= \gamma_{j,t-1} \cos\left(\frac{\pi j}{J/2}\right) + \gamma_{j,t-1}^* \sin\left(\frac{\pi j}{J/2}\right) + \omega_{j,t} \\ \gamma_{j,t}^* &= \gamma_{j,t-1}^* \cos\left(\frac{\pi j}{J/2}\right) - \gamma_{j,t-1} \sin\left(\frac{\pi j}{J/2}\right) + \omega_{j,t}^* \\ \omega_{j,t} &\simeq N(0, \sigma_\omega^2) \quad \omega_{j,t}^* \simeq N(0, \sigma_\omega^2), \quad j = 1, \ldots, J/2 \end{aligned}$$

For monthly time series, $J = 12$. Finally, ϵ_t in (2) denotes the irregular component, which contains the unexplained variation of the population parameter and is modelled as a white noise process, i.e. $\epsilon_t \simeq N(0, \sigma_\epsilon^2)$.

The systematic differences between the subsequent waves, i.e. the RGB, are modelled in (1) with Λ_t. The absolute bias in the monthly labour force figures cannot be estimated from the sample data only. Therefore, additional restrictions for the elements of Λ_t are required to identify the model. Here it is assumed that an unbiased estimate for θ_t is obtained with the first wave, i.e. $\hat{y}_t^{[1]}$. This implies that the first component of Λ_t equals zero. The other elements of Λ_t measure the time dependent differences with respect to the first wave and are modelled as random walks. As a result, it follows that

$$\begin{aligned} \lambda_t^{[1]} &= 0, \;\; \lambda_t^{[j]} = \lambda_{t-1}^{[j]} + \eta_{\lambda,t}^{[j]}, \;\; j = 2,3,4,5, & (3) \\ \eta_{\lambda,t}^{[j]} &\simeq N(0, \sigma_\lambda^2). & (4) \end{aligned}$$

Note that the disturbance terms of the random walks for the RGB in waves 2, 3, 4 and 5 share the same variance component σ_λ^2.

The discontinuities induced by the redesigns in 2010 and 2012 are modelled with the third and fourth term in (1). The diagonal matrix $\Delta_t^{[i]}$ contains five intervention variables:

$$\delta_t^{[i,j]} = \begin{cases} 0 & if \quad t < T_{R_i}^{[j]} \\ 1 & if \quad t \geq T_{R_i}^{[j]} \end{cases}, \qquad (5)$$

where $T_{R_i}^{[j]}$ denotes the moment that wave j changes from the old to the new survey design in redesign $i = 1$ in 2010 or $i = 2$ in 2012. Under the assumption that (2) correctly models the evolution of the population variable, the regression coefficients in $\mathrm{B}^{[i]}$ can be interpreted as the systematic effects of the redesign on the level of the series observed in the five waves. The intervention approach with state-space models was originally proposed by Harvey and Durbin (1986) to estimate the effect of seat belt legislation on British road casualties. With level intervention (5), it is assumed that the redesign only has a systematic effect on the level of the series. Alternative interventions, e.g. for the slope or the seasonal components are also possible, see Durbin and Koopman (2012), Ch. 3. A redesign might not only affect the point estimates, but also the variance of the GREG estimates. This issue is discussed under the time series model for the survey errors.

The last component in (1) is a time series model for the survey errors. The direct estimates for the design variances of the survey errors can be estimated from the micro data and are used as a-priori available parameters in the time series model. To account for heteroscedasticity in the sampling errors the following survey error model is proposed $e_t^{[j]} = k_t^{[j]} \tilde{e}_t^{[j]}$ where $k_t^{[j]} = \sqrt{\widehat{var}(\hat{y}_t^{[j]})}$, with $\widehat{var}(\hat{y}_t^{[j]})$ the estimated variance of the GREG estimator $\hat{y}_t^{[j]}$. Choosing the survey errors proportional to the standard error of the GREG estimators allows for non-homogeneous variance in the survey errors, that arise e.g. due to the gradually changing sample sizes over time.

The sample of the first wave has no sample overlap with waves observed in the past. Consequently, the survey errors of the first wave, $e_t^{[1]}$, are not correlated with survey errors in the past. It is, therefore, assumed that $\tilde{e}_t^{[1]}$ is white noise, i.e. $\tilde{e}_t^{[1]} \simeq N(0, \sigma_{e_1}^2)$. As a result, the variance of the survey error equals $var(e_t^{[1]}) = (k_t^{[1]})^2 \sigma_{e_1}^2$, which is approximately equal to the direct estimate of the variance of the GREG estimate for the first wave if the maximum likelihood (ML) estimate for $\sigma_{e_1}^2$ is close to one.

The survey errors of the second, third, fourth and fifth wave are correlated with survey errors of preceding periods. The autocorrelations between the survey errors of the subsequent waves are estimated from the survey data, using the approach proposed by Pfeffermann et al. (1998). In this application, it appears that the autocorrelation structure for the second, third, fourth and fifth wave can be modelled conveniently with an AR(1) model, van den Brakel and Krieg (2009). Therefore, it is assumed that $\tilde{e}_t^{[j]} = \rho \tilde{e}_{t-3}^{[j-1]} + v_t^{[j]}$, with ρ the first order autocorrelation coefficient, and $v_t^{[j]} \simeq N(0, \sigma_{e_j}^2)$ for $j = 2, 3, 4, 5$. Since $\tilde{e}_t^{[j]}$ is an AR(1) process, $var(e_t^{[j]}) = (k_t^{[j]})^2 \sigma_{e_j}^2 / (1 - \rho^2)$. As a result, $var(e_t^{[j]})$ is approximately equal to $\widehat{var}(\hat{y}_t^{[j]})$ provided that the ML estimates for $\sigma_{e_j}^2$ are close to $(1 - \rho^2)$.

The survey redesign in 2010 and 2012 might affect the variance of the GREG estimates. Systematic differences in these variances are automatically taken into account, since they are used as a-priori information in the time series model for the survey error. An alternative possibility would be to allow for different values for $\sigma_{e_j}^2$ for the periods with different survey design, which can be interpreted as an intervention on the variance hyperparameter of the survey error.

The general way to proceed is to express the model in the so-called state-space representation and apply the Kalman filter to obtain optimal estimates for the state variables, see e.g. Durbin and Koopman (2012). It is assumed that the disturbances are normally distributed. Under this assumption, the Kalman filter gives optimal estimates for the state vector and the signals. Estimates for state variables for period t based on the information available up to and including period t are referred to as the filtered estimates. The filtered estimates of past state vectors can be updated if new data become available. This procedure is referred to as smoothing and results in smoothed estimates that are based on the completely observed time series. In this application, interest is mainly focused on the filtered estimates, since they are based on the complete set of information that would be available in the regular production process to produce a model-based estimate for month t.

The analysis is conducted with software developed in OxMetrics in combination with the subroutines of SsfPack 3.0, see Doornik (2009) and Koopman et al. (2008). All state variables are non-stationary with the exception of the survey errors. The non-stationary variables are initialised with a diffuse prior, i.e. the expectation of the initial states are equal to zero and the initial covariance matrix of the states is diagonal with large diagonal elements. The survey errors are stationary and therefore initialised with a proper prior. The initial values for the survey errors are equal to zero and the covariance matrix is available from the aforementioned model for the survey errors. In Ssfpack 3.0, an exact diffuse log-likelihood function is obtained with the procedure proposed by Koopman (1997).

4 Redesign of the LFS

4.1 Quantifying discontinuities

As already mentioned in the introduction, Statistics Netherlands is planning a third redesign of the LFS that will be implemented in 2021. Important planned changes are a transition from a household sample to a personal sample and a revision of the questionnaire and the fieldwork strategy. Such changes generally have a systematic effect on the results of a sample survey. In this transition the software that support the data collection and processing will also be updated because a new version of Blaise will be implemented. It is, however, anticipated that this does not contribute to the discontinuities. To avoid confounding real developments with systematic effects induced by the redesign, it is important to quantify these discontinuities and to account for these effects in the time series model, as explained in Section 3. There are several methods that can be used to quantify discontinuities, see van den Brakel et al. (2020) for an overview. In this paper we investigate two approaches and the combination of them.

The first approach is to collect data under the old and new survey designs alongside each other for some period. This is referred to as parallel data collection or parallel run. A parallel run is preferably designed as a randomised experiment, where the sampling units from a probability sample are randomised over the current and alternative survey designs such that the subsamples can be considered as the treatment groups in an experiment. Embedding randomised experiments in probability samples generally increases the validity of the results. The theory of experimental design focuses on establishing the causality between treatments and effects. Selecting the experimental units randomly using probability sampling enables the generalization of conclusions of an experiment to larger target populations. This is particularly important if experiments are conducted to obtain quantitative insights into the effect of a new survey process on the outcomes of a repeated survey. In such cases, a design-based inference procedure for estimating discontinuities in sample estimates for unknown population parameters naturally fits with the purpose of probability sampling to generalize conclusions to larger target populations. A design-based inference procedure for this type of experiments is proposed by van den Brakel and Renssen (1998, 2005); van den Brakel (2008, 2013); Chipperfield and Bell (2010).

A strong point of a parallel data collection is the low risk level to the regular publications during the changeover to the new design. This approach can avoid the risk of a period without data for regular publication should the new survey design turn out to be a failure. Through a well-designed experiment the risk of failing to detect a discontinuity is minimized since the design of an experiment gives full control over the minimum detectable difference at a pre-specified significance and power level. A further major advantage of parallel data collection is that it facilitates the production of timely discontinuity estimates. Final estimates for the discontinuity can be made directly after finishing the fieldwork.

A disadvantage of a parallel run is that extra cost is required for additional data collection. Obtaining sufficiently precise estimates for the discontinuities often requires large sample sizes. Designing and conducting an experiment for parallel data collection that accurately measures the discontinuities due to the changeover also significantly increases the complexity of the fieldwork operation.

The second approach to quantify discontinuities is by fitting a structural time series model, containing intervention variables to account for discontinuities, as explained in Section 3 for the two redesigns in 2010 and 2012. A major advantage of the time series approach is that no additional data collection is required, which makes this approach very cost-effective. In addition, the complications of embedding a parallel run in the survey fieldwork are avoided. Another advantage of the time series modelling approach is that all available data under both the old and the new survey designs are used, since the entire observed series is used to estimate discontinuities. Disadvantages of this approach during the first periods after the change-over are that the estimates for the discontinuities are very uncertain and revised each time a new observation becomes available. This method also relies on the assumption that the time series components for the population parameter model are correctly specified. In that case all deviations from the expected evolution are interpreted as the discontinuity induced by the redesign. If, however, the transition coincides with a turning point of the population parameter, it can be expected that a part of the real evolution will be incorrectly interpreted as a discontinuity.

The time series approach, however, has the flexibility to combine information in the entire series with the information obtained with a parallel data collection. This can be done by using the sample estimates for discontinuities and their variances obtained with a parallel data collection for an exact initialisation of the regression coefficient of the intervention variable in the Kalman filter. In this way, the information obtained with a parallel run of insufficient sample size, can be further improved with the information from the entire time series observed before and after the parallel run.

4.2 Transition plan

To account for the discontinuities of the planned redesign in 2021 the time series model used for the production of monthly labour force figures, model (1) will be extended with a third intervention component, i.e.

$$\hat{\mathbf{y}}_t = \mathbf{j}_{[5]}\theta_t + \Lambda_t + \Delta_t^{[1]}\mathrm{B}_t^{[1]} + \Delta_t^{[2]}\mathrm{B}_t^{[2]} + \Delta_t^{[3]}\mathrm{B}^{[3]} + \mathbf{e}_t, \tag{6}$$

were $\Delta_t^{[3]}$ is defined in (5) and $\mathrm{B}^{[3]}$ are the regression coefficients which can be interpreted as the discontinuities in the different waves, due to the upcoming redesign. Contrary to the redesigns in 2010 and 2012, it is planned to do a parallel run in the first wave before the start of the transition. The samples in the first wave under the new design are not used in the production and these samples will not be reinterviewed for times. This is the reason that all regression coefficients in $\mathrm{B}^{[3]}$ are time invariant.

As explained in Section 3, the population parameter estimates in the time series model are benchmarked to the level of the series observed in the first wave, since it is assumed that the RGB in the first wave equals zero. It is therefore crucial that the first wave is measured as accurately as possible, including possible discontinuities due to a redesign. Therefore, it is decided to allocate the available budget for a parallel run for the first wave exclusively, and use the intervention approach for the remaining waves. This strategy is inline with the transition chosen for the redesigns in 2010 and 2012 (van den Brakel and Krieg, 2015).

Since the available budget for parallel data collection is limited, it is also decide that the primary purpose of the parallel run is to estimate the net effect of the change as precise as possible and not to disentangle the contributions of the different factors that changed in the redesign. This choice allows to design the parallel run as a two-sample experiment where the current and new survey designs are compared. If the purpose of the parallel run is to explain the individual contributions of the factors that changed in the redesign, then a factorial design is required, which means that the levels of all factors are combined in the experiment. This is at the cost of a reduced power for estimating the discontinuity of the new design, which can be seen by noticing that the overall discontinuity is one of the interactions of a factorial setup, namely the contrast between the subsample where all factors are on the level of the current survey design and the subsample where all factors are on the level of the new survey design.

Thanks to a grant from the European Commission, there is budget to conduct the first wave under the old and new design in parallel on the full sample size for a period of three months. Based on above considerations, the transition starts with a parallel run of three months in the first wave in October 2020. The three samples under the new design obtained during the parallel run won't go to the follow up waves. So the actual transition starts in January 2021 in the first wave. From that date on the new design is gradually phased in, which means that the second wave changes from the old to the new design in April 2021. The third wave changes from the old to the new design in July 2021, the fourth wave in October 2021 and the fifth wave in January 2022.

With the parallel run a first direct estimate for the discontinuity in the first wave is obtained. Let $\hat{y}_t^{[1,N]}$ denote the GREG estimate under the new design in month t observed in the first wave and $\hat{y}_t^{[1,O]}$ the GREG estimate under the old design in the first wave. Then $\hat{\Xi}_t = \hat{y}_t^{[1,N]} - \hat{y}_t^{[1,O]}$ denote the GREG estimate for the discontinuity in the first wave in month t, and $\hat{\hat{\Xi}} = (1/T_p)\sum_{t=1}^{T_p} \hat{\Xi}_t$, the direct estimate for the discontinuity obtained with the parallel run, with T_p the length of the parallel run in months. It is understood that in the expression for $\hat{\hat{\Xi}}$, $t = 1$ refers to the first month of the parallel run. The variance of the direct estimate of the discontinuity is obtained by $\hat{var}(\hat{\hat{\Xi}}) = \frac{1}{T_p^2}\sum_{t=1}^{T_p}(\hat{var}(\hat{y}_t^{[1,N]}) + \hat{var}(\hat{y}_t^{[1,O]}))$, see van den Brakel and Renssen (2005) for expressions of $\hat{var}(\hat{y}_t^{[1,N]})$ and $\hat{var}(\hat{y}_t^{[1,O]})$ under embedded randomized experimental designs. Based on variance calculations, standard errors for $\hat{\hat{\Xi}}$ are calculated if the first wave is conducted in parallel under the old and new design for a period of 3 and 6 and 12 months for the employed, unemployed labour force and total labour force at the national level and a breakdown in six domains used for the monthly publication. Results are presented in Table 4.1.

Parameter	National	M15-24	W15-24	M25-44	W25-44	M45-74	W45-74
	$T_p = 12$ months						
Unempl.	12	4	4	6.5	5.5	5.5	5
Empl.	22.5	7.5	7.5	9	11	11	12.5
	$T_p = 6$ months						
Unempl.	17	6	6	9	8	8	7
Empl.	32	11	11	13	16	16	18
	$T_p = 3$ months						
Unempl.	24	8.5	8.5	13	11	11	10
Empl.	45	15.5	15.5	18	22.5	22.5	25.5

Table 4.1 Standard errors of the GREG estimates of discontinuities ($\times$ 1000) for parallel runs of 12, 6 and 3 months.

It is anticipated that there is budget for a parallel run with a length of three months. The estimates obtained from a parallel run can be further improved in the time series modeling approach. To this end, the initial estimates obtained with the parallel run are used for an exact initialization of the state variable that models the discontinuity for the first wave. The information observed before and after the change over to the new design is used in the time series model to further improve the estimate for the discontinuity. In Section 5, a simulation is proposed to evaluate at which precision discontinuities in the monthly figures of the Dutch LFS can be estimated under different scenarios, e.g. the time intervention approach of the time series model without parallel data collection or the time series model in combination with a parallel run of different lengths. In addition, it is evaluated at which time the official publications can change from the level under the old approach to the level of the new approach.

5 Simulation

To simulate with which precision discontinuities can be estimated with the proposed time series modeling approach, time series for the five waves as observed with the LFS are generated from the assumed distribution under model (1). This is done as follows:

- Fit model (1) to the five time series observed with the LFS to obtain the maximum likelihood estimates for the hyperparameters for the disturbance terms of the trend (σ_η^2), the seasonal component (σ_ω^2), the RGB (σ_λ^2), and the sampling error ($\sigma_{e_j}^2$)
- The values for these hyperparameters define the distribution of the simplified model
$$\hat{\mathbf{y}}_t = \mathbf{j}_{[5]}\theta_t + \Lambda_t + \mathbf{e}_t. \tag{7}$$
- Use (7) to generate 1000 five dimensional time series replicates that behave similar to the series initially obtained with the LFS. The length of the simulates series is 18 years. The transition starts at month 120, which is January of the tenth year.

Subsequently discontinuities of different sizes are added to the five waves of the time series. Under different scenarios, different values for $\mathrm{B} = (\beta^{[1]}, \beta^{[2]}, \beta^{[3]}, \beta^{[4]}, \beta^{[5]})^t$. Discontinuity $\beta^{[1]}$ is added to the first wave in period $t = T_R$, which denotes the start of the transition. In these simulations $T_R = 120$. Then discontinuity $\beta^{[2]}$ is added to the second wave in period $t = T_R + 3$, discontinuity $\beta^{[3]}$ is added to the third wave in period $t = T_R + 6$, etc. More generally, replicates of time series with discontinuities are obtained as:

$$\hat{\mathbf{y}}_t^* = \hat{\mathbf{y}}_t + \Delta_t \mathrm{B},$$

with

$$\delta_t^{[i,j]} = \begin{cases} 0 & if \quad t < T_R + (j-1)\times 3 \\ 1 & if \quad t \geq T_R + (j-1)\times 3 \end{cases}.$$

In a next step a time series model with an intervention component is applied to the generated series with the discontinuities, i.e.

$$\hat{\mathbf{y}}_t^* = \mathbf{j}_{[5]}\theta_t + \Lambda_t + \Delta_t\mathrm{B} + \mathbf{e}_t, \tag{8}$$

Subsequently it is analyzed with which accuracy the discontinuities B, and in particular $\beta^{[1]}$ for the first wave are recovered from the estimation results under (8). In a similar way the impact of the redesign on the bias and variance of trend L_t will be analyzed.

For this simulation, also replicates for the parallel run are needed. It is therefore assumed that the direct estimates of the discontinuities obtained with the parallel run follow a normal distribution with an expected value that is equal to the assumed discontinuity and a variance based on the standard errors reported in Table 4.1. As a result, replicates for the exact initialization of $\beta^{[1]}$ in the Kalman filter are obtained by drawing resamples from $\mathcal{N}(\beta^{[1]}, v\hat{a}r(\hat{\hat{\Xi}}))$, where $\beta^{[1]}$ is a fixed value defined on the chosen scenario's of the simulation and $v\hat{a}r(\hat{\hat{\Xi}})$ depends on the variable and length of the parallel run as specified in Table 4.1.

6 Results

For the unemployed and employed labour force at the national level, simulations are conducted for three different settings, which are specified in Table 6.1. The first scenario assumes positive discontinuities, the second scenario negative discontinuities and under the third scenario, there are no discontinuities. For each scenario a simulation with 1000 replicates is conducted.

Scenario	$\beta^{[1]}$ wave 1	$\beta^{[2]}$ wave 2	$\beta^{[3]}$ wave 3	$\beta^{[4]}$ wave 4	$\beta^{[5]}$ wave 5
Unemployd labour force national level					
1	50000	45000	40000	35000	30000
2	-50000	-45000	-40000	-35000	-30000
3	0	0	0	0	0
Employd labour force national level					
1	75000	80000	85000	90000	95000
1	-75000	-80000	-85000	-90000	-95000
3	0	0	0	0	0

Table 6.1 Values for the discontinuities used in three different scenarios in the simulations for the transitions with the parallel runs of 12, 6 and 3 months as well as a transition without a parallel run

Let $\beta^{[j]}_{t|t,r}$ and $se(\beta^{[j]}_{t|t,r})$ denote the filtered estimate for the discontinuity of wave j at period t and its standard error in replicate r of the simulation, $r = 1, \dots, R$ and R the total number of replicates (1000). Furthermore,

$$\bar{\beta}^{[j]}_{t|t} = \frac{1}{R}\sum_{r=1}^{R}\beta^{[j]}_{t|t,r}, \quad \bar{se}(\beta^{[j]}_{t|t}) = \frac{1}{R}\sum_{r=1}^{R} se(\beta^{[j]}_{t|t,r}) \tag{9}$$

are the means of the filtered estimates for discontinuity and their standard errors over the R replicates in the simulation. Although the $\beta^{[j]}$ are constant, the filtered estimates will change over time because they are updated by the Kalman filter each time a new observation becomes available. Particularly directly after the change-over to the new design the updates $\beta^{[j]}_{t|t}$ from period to period will be substantial in the case of no parallel run or a small parallel run.

In Figure 6.1, the $\bar{\beta}^{[1]}_{t|t}$ and $\bar{se}(\beta^{[1]}_{t|t})$ are plotted for the first 24 months after the change-over for the first wave of the total unemployed labour force at the national level under three different scenario's (see Table 6.1). This is done for a transition without a parallel run (PR0), and a parallel run of 3 months (PR3), 6 months (PR6) and 12 months (PR12). In the case of no parallel run, the initialization of $\beta^{[j]}$ in the Kalman filter is diffuse. In the case of a parallel run the initialization is exact, using a draw from $\mathcal{N}(\beta^{[1]}, \hat{var}(\hat{\bar{\Xi}}))$, where $\beta^{[1]}$ as explained in Section 5 and a variance that depends on the length of the parallel run as specified in Table 4.1.

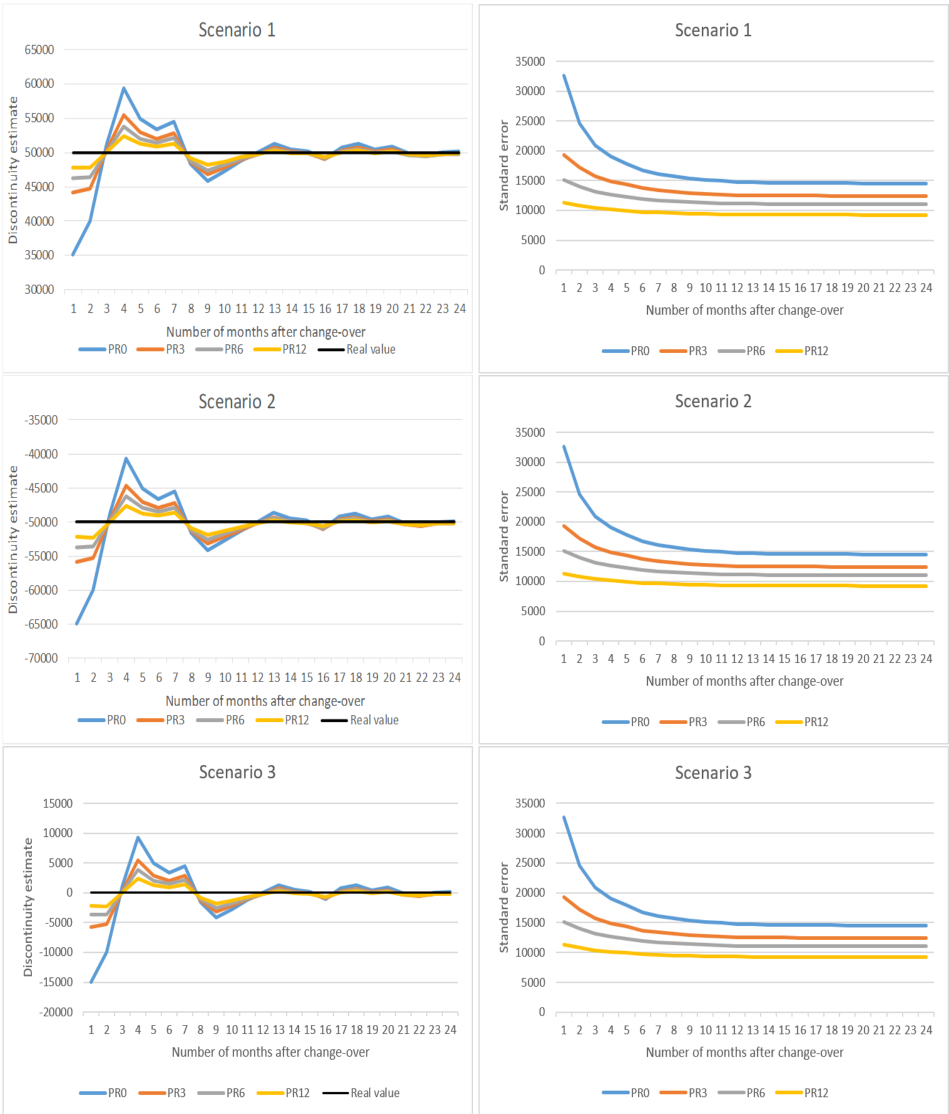

Figure 6.1 Simulated discontinuities and their standard errors for the unemployed labour force at national level in the first wave for different scenarios as specified in Table 6.1

On average the estimates for the discontinuities converge to their real values $\beta^{[j]}$. It takes about a year for a more or less stable estimate is obtained, regardless the length of the parallel run. In case of a parallel run the size of the revisions is smaller compared to change-over without a parallel run. The size of the revisions also decreases if the length of the parallel run increases. Similarly, the standard errors are the largest in case of a change-over without a parallel run, eventually these stabilize at a value of 14500. With a parallel run of three, six and twelve months, the standard errors stabilize at 12500, 11000, and 9000, respectively. The figure also illustrates that the differences between the filtered discontinuities and the real values as well as the standard errors are similar for the different scenarios. In other words, the precision of the filtered discontinuities does not depend on the assumed real values for $\beta^{[j]}$.

In Figure 6.2 similar plots for the $\bar{\beta}_{t|t}^{[1]}$ and $\bar{s}e(\beta_{t|t}^{[1]})$ are made for the employed labour force at the national level. Since the results are similar for the three different scenario's, the figures are only presented for Scenario 1. Results are similar as for the unemployed labour force. Point estimates are more or less stable after one year. Revisions and standard errors are the largest in the case of a transition without a parallel run and the standard errors stabilize at a level of 26000. With a parallel run of three, six and twelve months, the standard errors stabilize at 22500, 20000, and 17000, respectively.

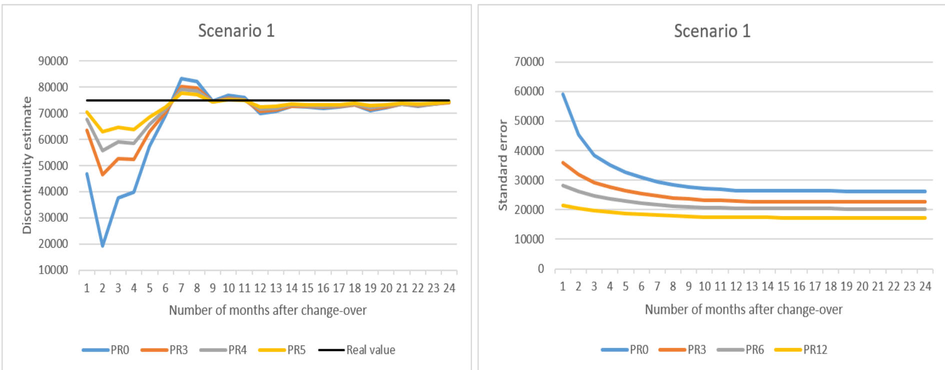

Figure 6.2 Simulated discontinuities and their standard errors for the employed labour force at national level in the first wave for Scenario 1 as specified in Table 6.1

Let $L^o_{t,r} = L_{t,r}$ denote the real trend under the old level, i.e. for the level under the approach for the transition to the new design, in the r-th replicate. Furthermore, $L^o_{t|t,r} = L_{t|t,r}$ and $se(L^o_{t|t,r}) = se(L_{t|t,r})$ denote the filtered estimate for $L^o_{t,r}$ at period t and its standard error for replicate r, $r = 1, \dots, R$. An impression of the bias in the trend estimates at the level of the old design before and after the transition to the new design is obtained with

$$\bar{\Delta}(L^o_{t|t}) = \frac{1}{R}\sum_{r=1}^{R}(L^o_{t,r} - L^o_{t|t,r}). \tag{10}$$

An indication of the precision of the trend under the old design is obtained with the mean of the $se(L^o_{t|t,r})$ over the R replications, say $\bar{s}e(L^o_{t|t})$, which is similarly defined as the means in (9). The mean over the absolute values of the difference between the real trend and the filtered trend is an other way to measure the increased uncertainty in the published trend, directly after the change-over to the new design, and is defined as:

$$A\bar{B}S[\Delta(L^o_{t|t})] = \frac{1}{R}\sum_{r=1}^{R}|L^o_{t,r} - L^o_{t|t,r}|. \tag{11}$$

The real trend under the new design in the r-th replicate is defined as $L^n_{t,r} = L_{t,r} + \beta^{[1]}$. Recall that the real discontinuity is equal for each replicate in the simulation. Furthermore, $L^n_{t|t,r} = L_{t|t,r} + \beta^{[1]}_{t|t,r}$ and $se(L^n_{t|t,r}) = se(L_{t|t,r} + \beta^{[1]}_{t|t,r})$ denote the filtered estimate for $L^n_{t,r}$ at period t and its standard error for replicate r, $r = 1, \dots, R$. Note that the standard error for $L^n_{t|t,r}$ accounts for the correlation between the filtered estimates for the trend and the discontinuity. An impression of the bias in the trend estimates at the level of the new design after the transition is obtained with

$$\bar{\Delta}(L^n_{t|t}) = \frac{1}{R}\sum_{r=1}^{R}[(L_{t,r} + \beta^{[1]}) - (L_{t|t,r} + \beta^{[1]}_{t|t,r})]. \tag{12}$$

An indication of the precision of the trend under the new design is obtained with the mean of the $se(L^n_{t|t,r})$ over the R replications, say $\bar{s}e(L^n_{t|t})$, which is similarly defined as the means in (9). The mean over the absolute values of the difference between the real trend and the filtered trend is also used to measure the increased uncertainty in the published trend under the new design, and is defined as:

$$A\bar{B}S[\Delta(L^n_{t|t})] = \frac{1}{R}\sum_{r=1}^{R}|L^n_{t,r} - L^n_{t|t,r}|. \tag{13}$$

The standard error of the trend under the old design ($se(L^o_{t|t})$) is compared with the standard error of the trend under the new design ($se(L^n_{t|t})$) for a transition without a parallel run and a parallel run with a length of three, six and twelve months in Figure 6.3 for the unemployed labour force at the national level. For the trend under the old design, the standard errors for the last thirteen months before the transition and the first 24 months after the transition are shown. For the trend under the new design, the standard errors only for the first 24 months after the transition are shown. Results are presented only for Scenario 1, since the results for Scenario 2 and 3 are exactly the same. Similar results are presented in Figure 6.4 for the employed labour force at the national level.

Figure 6.3 Standard errors of the simulated trends of Unemployed labour force at the national level under the old and new designs for a transition without a parallel run (PR0) and a parallel run with a length of 3 months (PR3), 6 months (PR6), and 12 months (PR12).

Figure 6.4 Standard errors of the simulated trends of Employed labour force at the national level under the old and new designs for a transition without a parallel run (PR0) and a parallel run with a length of 3 months (PR3), 6 months (PR6), and 12 months (PR12).

The Figures 6.3 and 6.4 illustrate some interesting results. The standard errors of the trend under the new design mainly benefit from a parallel run in the first months, directly after the change-over. After about 10 to 12 months the standard errors for the trend under the new design converge to the same level, regardless a transitions with or without a parallel run and regardless the length of the parallel run. The standard errors of the trend under the old design gradually increase after the change-over during a period of about 16 months. The additional amount of uncertainty added to the level of the trend under the old design is the highest under a transition without a parallel run and decreases with the size of the parallel run. The interpretation of this result is that the trend at the new level is the sum of the trend and the estimated discontinuity. After the change-over the level of this sum can be estimated better than their separate values. Therefore the uncertainty of the trend at the level of the old design is larger than the trend at the level of the new design and eventually the standard error of the trend under the old design benefit more from a parallel run than the trend under the new design.

After a period of 5 months the standard error of the trend under the new design drop below the standard error of the trend under the old design, regardless whether the transition is done with or without a parallel run. Also the length of the parallel run does not have an effect on this. If, however, the trend under the new design is backcasted for the period before the transition, the uncertainty for the trend for this period will depend more strongly whether a parallel run has been conducted and the size of this parallel run, since the uncertainty of backcasted trends for this period directly depend on the standard errors of the estimated discontinuities as illustrated in Figures 6.1 and 6.2.

After the start of the change-over, the standard errors of the trend under the old design and to a lesser extend also the trend under the new design steadily increase during a period of about 16 months. This is the result of activating a new intervention variable into the time series model, each quarter when the new design first enters another follow up wave. This results in a steady increasing amount of uncertainty, in particular for the level of the trend under the old design.

The mean of the bias or difference between the real and filtered trend under the old design, defined in (10), and the new design, defined in (12), is plotted in Figure 6.5 for the Unemployed labour force at the national level and in Figure 6.6 for the Employed labour force at the national level. Results are shown for a transition without a parallel run and a parallel run with a length of 3 months, 6 month and 12 months. Results are shown for Scenario 1 only and it is noted that the results are exactly equal for Scenarios 2 and 3. In all the situations, the bias fluctuates around zero. Directly after the change-over the bias in the trend under the new design is large compared to the trend under the old design. It takes on average three to five months before the bias in the trend of the new design is comparable to the bias in the trend of the old design. A parallel run mainly decreases the bias in the trend under the new design, in the first three to five months after the transition to the new design. The longer the parallel run the smaller the bias in this period.

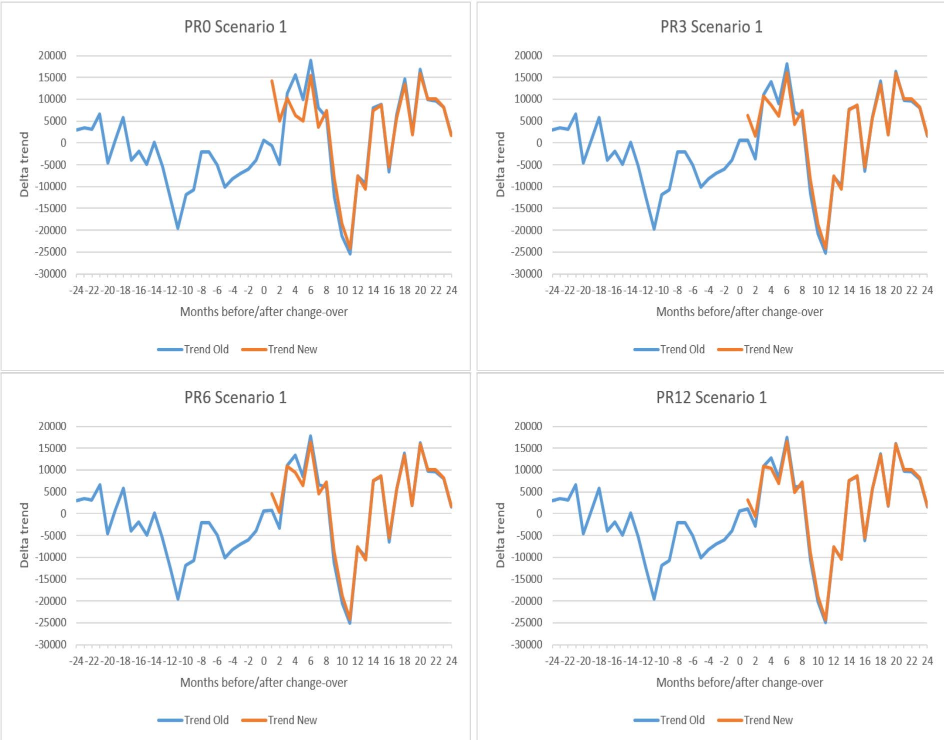

Figure 6.5 Mean bias in the simulated trends of Unemployed labour force at the national level under the old and new design for a transition without a parallel run (PR0) and a parallel run with a length of 3 months (PR3), 6 months (PR6), and 12 months (PR12).

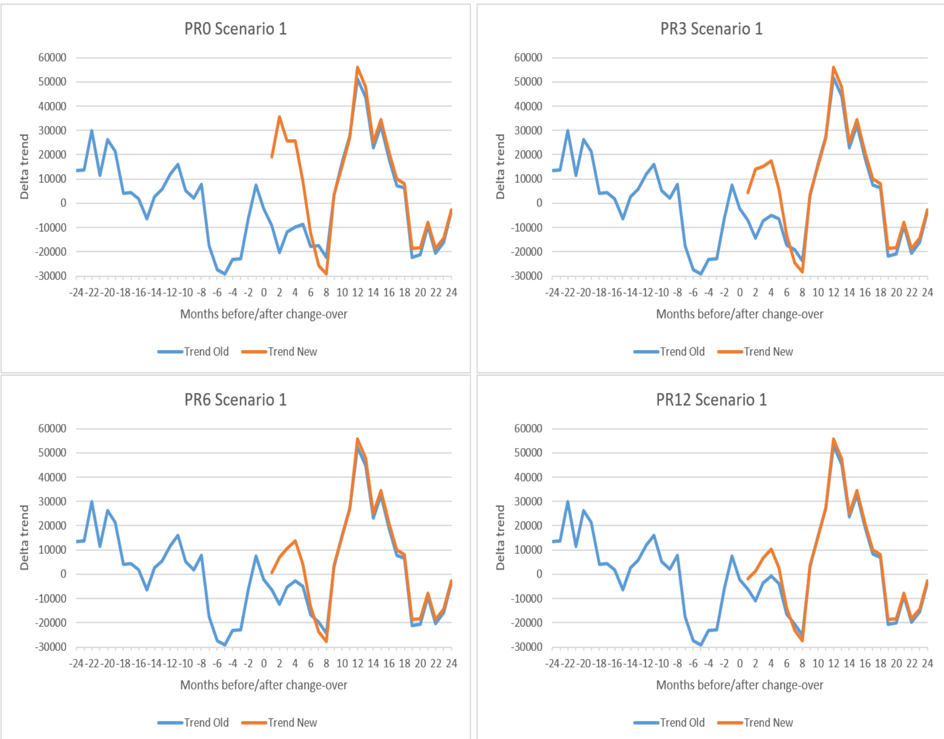

Figure 6.6 Mean bias in the simulated trends of Employed labour force at the national level under the old and new design for a transition without a parallel run (PR0) and a parallel run with a length of 3 months (PR3), 6 months (PR6), and 12 months (PR12).

The mean of the absolute difference between the real and filtered trend under the old design, defined in (11), and under the new design, defined in (13), is plotted in Figure 6.7 for the Unemployed labour force at the national level and in Figure 6.8 for the Employed labour force at the national level. As in the previous figures, results are shown for a transition without a parallel run and a parallel run with a length of 3 months, 6 month and 12 months and for Scenario 1 only, since the results for Scenario 2 and 3 are exactly equal. As observed for the bias, the parallel run and the length of it only effect the absolute difference in the trend under the new design in the first few months after the change-over to the new design.

Based on the available budget it is decided to conduct a parallel run of three months in the last quarter of 2020. The transition starts in January 2021. Table 6.2 specifies the expected standard errors of the estimated discontinuities for Unemployment at the national level and the six publication domains for gender by age for a transition without a parallel run and the planned transition with a parallel run of three months. Results are given for the first three, the sixth, the ninth, the twelved and the twenty-fourth month after the start of the change-over. This gives an expression of the accuracy with which discontinuities can be estimated under the planned transition.

		Months after start transition						
Domain	PRx	M1	M2	M3	M6	M9	M12	M24
National	PR0	32600	24500	20800	16700	15400	14700	14500
	PR3	19300	17200	15700	13700	12900	12600	12400
Men 15-24	PR0	13200	9600	8200	6300	5600	5400	5200
	PR3	7200	6400	5900	5100	4700	4500	4400
Women 15-24	PR0	13300	9700	8200	6400	5500	5100	4800
	PR3	7200	6500	6000	5200	4900	4400	4300
Men 25-44	PR0	15400	11500	9500	6800	6000	5600	5400
	PR3	9900	8600	7700	6100	5400	5100	5000
Women 25-44	PR0	12800	10000	8700	6900	6200	5900	5800
	PR3	8300	7400	6800	5800	5400	5200	5100
Men 45-74	PR0	14800	11200	9600	7200	6500	6200	6100
	PR3	8800	7800	7200	6000	5600	5400	5300
Women 45-74	PR0	12900	9200	7800	6200	5500	5300	5200
	PR3	7900	6800	6200	5300	4800	4700	4600

Table 6.2 Expected standard errors for the estimated discontinuities for Unemployed labour force at the national level and a breakdown in six publication domains for a transition without a parallel run and the planned transition with a parallel run of three months.

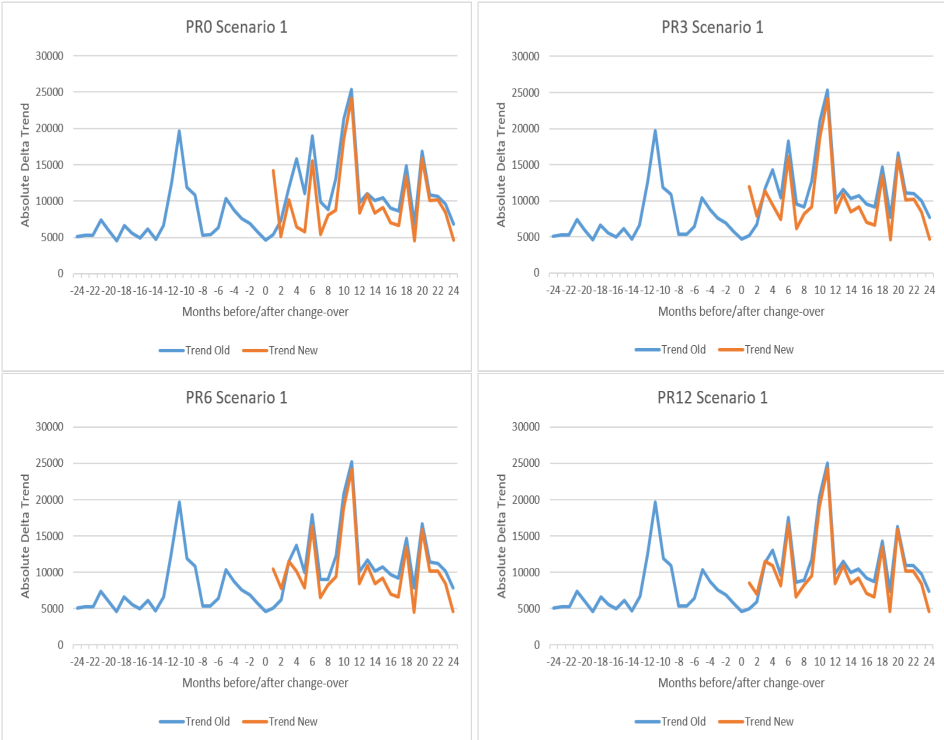

Figure 6.7 Mean absolute difference between real and filtered trends of Unemployed labour force at the national level under the old and new design for a transition without a parallel run (PR0) and a parallel run with a length of 3 months (PR3), 6 months (PR6), and 12 months (PR12).

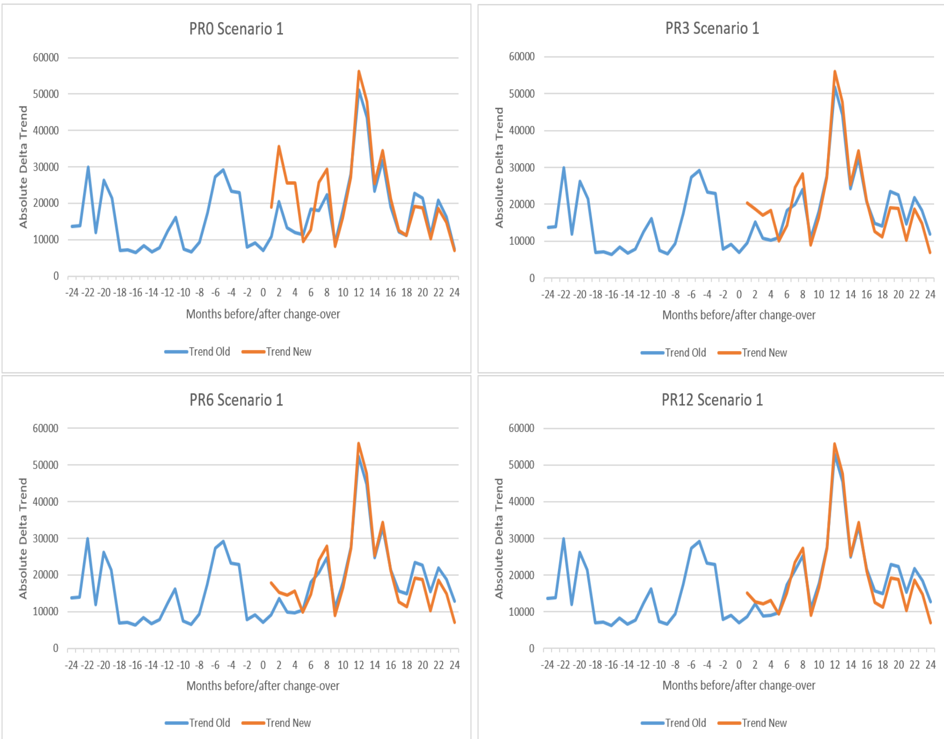

Figure 6.8 Mean absolute difference between real and filtered trends of Employed labour force at the national level under the old and new design for a transition without a parallel run (PR0) and a parallel run with a length of 3 months (PR3), 6 months (PR6), and 12 months (PR12).

7 Discussion

Monthly figures about the Dutch labour force are based on a structural time series model. This estimation approach solves multiple problems. First, monthly direct sample estimates are too volatile due to the relative small monthly sample size. The time series model is used to produce more precise and stable estimates for the monthly labour force figures by increasing the effective sample size of the last period with the sample information that is observed in the previous periods. From this perspective the time series model is used as a form of small area estimation. Second, the model accounts for rotation group bias, by benchmarking the model estimates to the level of the first wave. As a result, the estimates based on the rotating panel, which is introduced in 2000, are comparable with the period before 2000, when the LFS was a cross-sectional survey. Third, the time series model accounts for serial autocorrelation in the sample errors, due to the partial sample overlap of the rotating panel design. Finally, the model accounts for discontinuities that arise due to two major redesigns of the survey process in 2010 and 2012. A consequence of making substantial modifications in a survey process is that this generally results in systematic differences in the outcomes of the survey, also known as discontinuities. They are the net result of changes in measurement and selection bias, because of introducing different questionnaires and field work strategies. In an ideal survey transition process, discontinuities due to the redesign are quantified to avoid that real period-to-period changes are confounded with the discontinuities induced by the redesign. In the previous two redesigns discontinuities are quantified by conducting the first wave in parallel with a sufficiently large sample, such that discontinuities in the first wave could be estimated with a direct estimator. For the follow up waves, no parallel data collection was conducted. Discontinuities in the follow up waves are estimated with a level intervention component in the time series model.

A third redesign of the LFS is planned in 2021. For this redesign, the two aforementioned approaches are considered. The main advantage of a parallel run is that in case of a sufficiently large sample size, the risk of disturbing the production of official statistics during the transition is low and that estimates for the discontinuities are timely available. The main drawbacks are the high costs for additional data collection and the complications of the fieldwork to conduct a well designed parallel run. The main advantage of estimating discontinuities using an intervention component in the time series model is that it is cost effective because no additional data collection is required and that the additional complications for the field work to conduct a well designed parallel run are avoided. The major drawback is the high risk of disturbing the publication of the monthly labour force figures during the transition period. Furthermore, the discontinuity estimates are highly uncertain in the first months after the change-over and are subject to large revisions.

For the upcoming redesign both approaches, parallel data collection and the intervention time series model, are combined. The advantage of the structural time series model is that it can combine the information from a parallel run with the information of the entirely observed time series before and after the parallel run. The information about the discontinuities obtained in a relative short parallel run is used in the time series model via an exact initialization of the regression coefficient of the intervention variable in the Kalman filter. The Kalman filter further improves the precision of the discontinuity estimates with the observations that become available after finalizing the parallel run.

In preparation of the redesign in 2021, an extensive simulation is conducted to establish at which precision discontinuities can be estimated, as well as the trend under the level of the old and new designs. Four different situations are distinguished; a transition without a parallel run and a transition where the first wave is conducted in parallel at a full sample size of respectively three, six and twelve months. For the Unemployed labour force at the national level, the reduction of the standard error is about 14% with a parallel run of three months compared to a transition without a parallel run. In the case of a parallel run of six months and twelve months, this reduction is 24% and 36% respectively. In the case of the employed labour force, the reduction of the standard error is about 14%, 23% and 34% for a parallel run of respectively three, six and twelve months. Note that these percentages refer the standard error reduction that is achieved after about 12 months. Directly after the change over tot he new design, the standard error reduction with a parallel run are larger.

The simulation illustrates that the parallel run has a strong effect on the precision of the trend at the level of the old design for the period after the change-over to the new design. The precision of the trend at the level of the new design is mainly influenced by a parallel run and its length in the first six months after the change-over. On the long run, the precision of the trend under the new design is hardly influenced by a parallel run for the period after the change-over. The precision of backcasted trends to level of the new design for the period before the change-over is directly effected by a parallel run and its length.

Finally, the simulation illustrates that after 5 to 6 months after the start of the change-over, the standard error of the trend at the level of the new design becomes smaller than the trend at the old level. It is therefore recommended that in the first four or five months after the start of the transition, the official monthly labour force figures can be published at the level of the old design. The publications can switch to the level of the new design in month five or six after the start of the transition.

Disclaimer

The views expressed in this paper are those of the authors and do not necessarily reflect the policy of Statistics Netherlands.

Reviewed by Sabine Krieg and Martijn Souren.

References

Bailar, B. (1975). The effects of rotation group bias on estimates from panel surveys. *Journal of the American Statistical Association 70*(349), 23–30.

Chipperfield, J. and P. Bell (2010). Embedded experiments in repeated and overlapping surveys. *Journal of the Royal Statistical Society, Series A 173*, 51–66.

Doornik, J. (2009). *An Object-oriented Matrix Programming Language Ox 6*. Timberlake Consultants Press.

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

Harvey, A. and J. Durbin (1986). The effects of seat belt legislation on british road casualties: a case study in structural time series modelling. *Journal of the Royal Statistical Society, A series 149*, 187–227.

Koopman, S. (1997). Exact initial kalman filtering and smoothing for non-stationary time series models. *Journal of the American Statistical Association 92*, 1630–1638.

Koopman, S., A. Shephard, and J. Doornik (2008). *Ssfpack 3.0: Statistical algorithms for models in state-space form*. Timberlake Consultants, Press London.

Pfeffermann, D. (1991). Estimation and seasonal adjustment of population means using data from repeated surveys. *Journal of Business & Economic Statistics*, 163–175.

Pfeffermann, D., M. Feder, and D. Signorelli (1998). Estimation of autocorrelations of survey errors with application to trend estimation in small areas. *Journal of Business & Economic Statistics*, 339–348.

Särndal, C.-E., B. Swensson, and J. Wretman (1992). *Model Assisted Survey Sampling*. Springer.

van den Brakel, J. and S. Krieg (2009). Structural time series modelling of the monthly unemployment rate in a rotating panel. *Survey Methodology 35*(2), 177–190.

van den Brakel, J. and S. Krieg (2015). Dealing with small sample sizes, rotation group bias and discontinuities in a rotating panel design. *Survey Methodology 41*(2), 267–296.

van den Brakel, J. and R. Roels (2010). Intervention analysis with state-space models to estimate discontinuities due to a survey redesign. *Annals of Applied Statistics 4*, 1105–1138.

van den Brakel, J., X. M. Zhang, and S.-M. Tam (2020). Measuring discontinuities in time series obtained with repeated sample surveys. *International Statisitcal Review published online*.

van den Brakel, J. A. (2008). Design-based analysis of experiments with applications in the dutch labour force survey. *Journal of the Royal Statistical Society, Series A 171*, 581–613.

van den Brakel, J. A. (2013). Design-based analysis of factorial designs embedded in probability samples. *Survey Methodology 39*, 323–349.

van den Brakel, J. A. and R. H. Renssen (1998). Design and analysis of experiments embedded in sample surveys. *Journal of Official Statistics 14*, 277–295.

van den Brakel, J. A. and R. H. Renssen (2005). Analysis of experiments embedded in complex sampling designs. *Survey Methodology 31*, 23–40.

Colophon

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
Statistics Netherlands, Grafimedia

Design
Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information