1351.1.55.160

Australian Bureau of Statistics

Research Paper

Exploration of state space modelling approaches for statistical impact measurement in ABS time series: The Labour Force Survey as a case study

Australia

2018

AUSTRALIAN BUREAU OF STATISTICS

Research paper

Exploration of state space modelling approaches for statistical impact measurement in ABS time series: The Labour Force Survey as a case study

Oksana Honchar and Cedric Wong

Methodology Division

Methodology Advisory Committee

14 June 2018, Canberra

ABS Catalogue No. 1351.1.55.160

Produced by the Australian Bureau of Statistics.

INQUIRIES

- For further information about these and related statistics, contact the National Information and Referral Service on 1300 135 070.

EXPLORATION OF STATE SPACE MODELLING APPROACHES FOR STATISTICAL IMPACT MEASUREMENT IN ABS TIME SERIES: THE LABOUR FORCE SURVEY AS A CASE STUDY

Oksana Honchar and Cedric Wong,
Methodology Division

QUESTIONS FOR THE COMMITTEE

1. Do you agree that the full model is a better choice than the other models? Is any improvement possible to the model?
2. Do you agree with the approach of standard error estimation for overall impact?
3. Can we improve the estimation of overall impact by considering bias-correction for the back-transformation?
4. Is ignoring correlation between outgoing (wave 8) and incoming (wave 1) rotation groups acceptable?
5. Is using constant sampling error variance across time and across waves a reasonable assumption?

CONTENTS

APPENDIXES

ABBREVIATIONS

ABS	Australian Bureau of Statistics
AC	Auto Correlation
AR	Auto Regressive
BLUE	Best Linear Unbiased Estimator
GREG	Generalised Regression
KF	Kalman Filter
LS	Level Shift
LFS	Labour Force Survey
MLE	Maximum Likelihood Estimation
MDI	Minimum Detectable Impact
PAC	Partial Auto Correlation
RG	Rotation Group
RGB	Rotation Group Bias
RSE	Relative Standard Error
SaE	Sampling Error
SIM	Statistical Impact Measurement
SSM	State Space Model
SE	Standard Error
STM	Structural Time Series Model

EXPLORATION OF STATE SPACE MODELLING APPROACHES FOR STATISTICAL IMPACT MEASUREMENT IN ABS TIME SERIES: THE LABOUR FORCE SURVEY AS A CASE STUDY

Oksana Honchar and Cedric Wong,
Methodology Division

ABSTRACT

The ABS is embarking on a transformation program, which includes a change in ABS systems (editing, imputation, estimation, etc.), a use of different collection modes and a change in sampling frames etc. Once this transformation is completed, it is expected to deliver positive changes to the production of official statistics. However, there is also a risk of introducing impacts on some ABS time series. The ABS is working on developing methods to measure, and where needed to adjust for any impacts on such time series.

This paper discusses the usage of state space modelling approach to measure statistical impacts under two scenarios: 1) having a small parallel collection and then incrementing (phasing-in) the new approach to the survey, 2) no parallel collection is available and a new approach is introduced gradually. We consider a few models in this study, including the difference model for parallel collection, phase-in model with and without Kalman filter initialisation using estimated impacts from parallel collection initially presented at the MAC May 2017 and finally, the full model that utilises both parallel collection and phase-in period information. The full model builds on feedback led by Professor Hyndman at the MAC May 2017. The models' performance is evaluated from the results of a simulation study for the Australian Labour Force Survey (LFS). Full model with parallel collection performs the best in terms of the power of detecting overall impacts when compared to the other models.

1. INTRODUCTION

The ABS is embarking on a transformation program, which includes a change in ABS systems (editing, imputation, estimation, etc.), a use of different collection modes and a change in sampling frames etc. Once this transformation is completed, it is expected to deliver positive changes to the production of official statistics. However, there is also a risk of introducing impacts on some ABS time series. The ABS is working on developing methods to measure, and where needed to adjust for any impacts on such time series.

The ABS uses a three stage approach to statistical impact management. The first stage involves stakeholder engagement, risk and project management, pre-testing, analysis, dissemination planning and other tools to manage statistical impacts from process changes. The second stage requires data collection from a trial or experiment and statistical modelling in order to estimate the magnitude of the impacts and to test their significance (impact measurement). Once an impact has been detected, there is a need for impact adjustments and efficient dissemination strategy (stage 3). This paper focuses on the second stage, which is called statistical impact measurement (SIM).

The choice of method for statistical impact measurement depends on a few factors such as the expected changes in the survey, availability of resources, the clients' requirement for the minimum detectable impact and in case of multiple changes, whether it is possible to measure the separated effect for each survey change or the accumulative effect of all changes. If the expected changes (such as changes of processing, imputation or estimation) do not affect raw unit-level data then impacts can be detected by running the old and new systems in parallel and then comparing estimates from the two systems. However, if changes are expected at the raw unit-level data (such as, change of survey mode) then impact measurement becomes more challenging. The most accurate impact estimates in this case can be obtained by using methods of embedded experiments that require a parallel collection. Under this approach, the old and new surveys are run in parallel for a period of time and this allows simultaneous comparison of estimates from the survey under the old and new approach. However, this approach is very costly and requires lots of resources and thus, it is often not feasible.

When parallel collection is not possible, there are other methods for SIM that can be loosely categorised into two groups: SIM methods with unit-level data and SIM methods with aggregated level estimates. The first group includes methods of matching (on covariates, propensity scores etc.) and unit-level modelling. The second group consists of time series modelling. In this paper, we will consider time series approach to SIM. We will make use of multivariate structural time series models in state space form. Different versions of this approach are considered in earlier studies by Pfeffermann (1991), Van Den Brakel (2009, 2010). Our initial research demonstrates that this approach can potentially provide accurate estimates of impacts using data from parallel collection and phase-in periods (Zhang *et. al.*, 2018). In this study, we propose a different state space model that has practical benefit as it can be used for both parallel collection and phase-in periods. In addition, it can measure impacts on the LFS series more precisely. The initial idea of this model was suggested by Professor Rob Hyndman following the MAC meeting in May 2017. We have extended his proposed model to take into account the sampling error structure of the LFS series and the parallel collection design.

We explore the state space modelling approach to measure impacts on Labour Force Survey (LFS) time series as permanent level shifts. Changes to LFS might introduce other types of impacts such as temporary effects and changes in seasonal pattern. However, we do not consider measuring such impacts in this paper because although state space models can potentially handle such impacts, it would take a long time for the model to estimate it. It is expected that the SIM model is capable of measuring impacts during a short period of time such as during parallel collection and phase-in period.

We are focusing on the measurement of impacts on the national LFS employment / unemployment time series. Impact measurement for lower aggregated time series such as employment / unemployment series by states, sex and age groups will be considered in the next study. Although the main interest here is to measure impacts on national estimates, impact measurement modelling has to be conducted at the wave (time-in-survey) level to take into account the fact that impacts are potentially time-in-survey sensitive. Note that impacts have to be detected on composite estimator series (the estimator used in the LFS). However, it is impossible to create time series at the wave level from the composite estimator. This is because composite estimator is a weighted average of 56 estimates from 8 rotation groups in the current and previous 6 months. Therefore, we fit a model to GREG estimator time series at the wave level and then combine detected impacts into an overall impact and adjust composite estimator time series by the size of detected impacts.

Section 2 provides a brief introduction of the characteristics of the current ABS LFS survey, possible future changes and basic structural time series model in state space form on real LFS data before the transformation. Section 3 describes the models that could be used for impact measurement in the LFS during parallel collection if it is available and phase-in periods. Section 4 presents an integrated approach for impact measurement with full model. Section 5 describes a data simulation process for model evaluation and method of estimating an overall impact and its standard error. Section 6 discusses performance of different models on simulated data. Finally, Section 7 gives a brief discussion of findings and concluding remarks.

All the calculations reported in this paper are carried out with programs written in the SSM procedure in SAS and the DLM package in R.

2. ABS LABOUR FORCE SURVEY (LFS)

2.1 LFS design and estimation

The Australian Labour Force Survey (LFS) is a monthly survey with approximately 26,000 households selected in the sample. This covers approximately 0.32% of the civilian population of Australia aged 15 years and over (ABS, 2018). A multi-stage survey design is used to select households in the sample with clustering or stratification of dwellings at each selection stage. Households that are selected for the LFS participate in the survey for eight consecutive months and then they are replaced by their next-door neighbour household. Thus the LFS sample comprises eight sub-samples (or rotation groups, RG hereafter) with each sub-sample remaining in the survey for eight months. Each month one rotation group "rotates out" and is replaced by a new group "rotating in". This is known as rotating panel design.

Strong overlap between samples in any two successive months (about seven eighths) induces a strong serial correlation in the sampling error. In fact, sampling error has a degree of positive correlation even between rotated-out and rotated-in groups because the replacement sample is selected from the same geographic areas as the outgoing one. This induces additional serial correlation in the sampling error even for non-overlapping outgoing and incoming RGs.

Traditionally, the first interview is generally conducted face-to-face and subsequent interviews are conducted by telephone. From December 2012 to April 2013, the ABS conducted a trial of online electronic data collection. Respondents in a single rotation group (i.e. one-eighth of the survey sample) were offered the option of self-completing their LFS questionnaire online instead of via a face-to-face or telephone interview. In May 2013, the ABS expanded the offer of online electronic collection to 50% of each new incoming rotation group. Since September 2013, online electronic collection has been offered to 100% of private dwellings in each incoming rotation group. Since April 2014, 100% of private dwellings across the sample have been offered online electronic collection. During the implementation period, the ABS investigated possible impacts of mode changes on LFS estimates using different statistical impact measurement techniques such as matching on covariates, propensity score matching, longitudinal analysis methods as well as state space modelling (later on) and no impacts on the LFS estimates were detected.

The estimation method in the LFS is composite estimation. By exploiting the high correlation between overlapping samples across months, the LFS composite estimator combines the current month data with previous months' data by applying different factors (called BLUE multipliers) based on the length of time a rotation group has been in the survey. After these factors are applied, seven months of data (current and previous six months) are weighted to align with known current month population benchmarks.

There are systematic differences in LFS estimates for the subsequent waves (month-in-survey) particularly for unemployment. The level of unemployment declines if the rotation group is in the survey for a longer period of time. In the literature, this is referred to as rotation group bias (RGB). This fact is well known in other countries, e.g. in the U.S. (Bailar, 1975), Canada (Kumar and Lee, 1983, Binder and Hidiroglou, 1988) and the Netherlands (Brakel and Krieg, 2009). The RGB has to be taken into account when impact measurement is conducted at the rotation group level due to e.g. the introduction of time-in-survey sensitive changes to the survey.

In some calendar months the LFS has supplementary surveys conducted in conjunction with the LFS. Some of labour-related supplementary surveys affect original labour force estimates. The identified impacts are adjusted by the ABS in seasonally adjusted series (but not in original series). There are also known cases when changes in supplementary survey program have additional temporary impact on the LFS estimates, e.g. in August 2014 when changes in supplementary survey program affected the seasonal pattern in employment series (see ABS, 2014). If the supplementary survey program changes further, then an impact of this change on the LFS series is almost certain.

2.2 Different types of transformative change

There is a broad range of potential future changes that are being explored for the LFS, as the ABS continues to modernise its statistical methodology. They have been conceptualised, for the purpose of this paper, as falling within a three stage model, according to the extent to which they represent a change from the current design of the survey, and the likely level of risk of a statistical impact.

- Changes that have been conceptualised as falling within a first stage are either regular updates to the survey design (e.g. a new sample design, which for 2018 includes the use of the Address Register) or have controllable impacts (e.g. estimation methodology changes). There is no need for SIM for those changes.
- Changes in a second stage would be those tied to major changes in the systems and infrastructure, and related processes. Impacts due to such changes are rather unlikely. For managing statistical risk purposes, parallel processing and pretesting with small size sample are recommended.
- Changes in a third stage would be those focused on the long-term sustainability of the LFS. They would involve a greater degree of change that could result in considerable changes in respondent behaviour – with likely mode and seasonality effects. The two such changes would be optimising the LFS questionnaire around e-collection as the primary mode for collection and in the placement and timing of supplementary surveys (except for a scenario where supplementary surveys were all removed entirely). Design around e-collection as the primary mode implies the potential for changes to the questions asked and the number of questions, respondent induction and follow-ups and therefore, there might be time-in-survey

sensitive effects. This implies the impacts could vary the longer the people stay in survey. There may be a uniform increase in the time-in-survey and it is expected that the drop off rate of e-collection across months would be much smaller than under the current approach. Both changes could have significant impacts on the LFS estimates, particularly changes in supplementary surveys program (see 2.1). Supplementary survey program changes would also be expected to be time-in-survey sensitive. We refer to this as "wave sensitive" in this paper.

2.3 Modelling LFS time series at the rotation group level

Let us first consider the structural time series model in state space form for employment / unemployment time series at the wave level before the ABS transformation. Assume $y_{i,t}$ is a GREG estimate of a main LFS variable such as the number of employed and the number of unemployed persons from the rotation group that is observed for the *i*th time in survey (*i=1, …, 8*) (refer to as wave *i* hereafter) in the current month *t*. Note that $y_{i,t}$ are obtained by using the GREG estimator with calibration of each rotation group estimates to the total known population benchmarks. We can use the following eight-dimensional multivariate structural time series model for modelling series $y_{i,t}$ [1]:

$$\mathbf{y_t} = \mathbf{I}_{[8\times8]}\left(T_t + S_t + I_t\right) + \mathbf{b} + \mathbf{e_t} \tag{2.1}$$

or in an expanded form:

$$\begin{bmatrix} y_{1,t} \\ \vdots \\ y_{8,t} \end{bmatrix} = \mathbf{I}_{[8\times8]}\left(T_t + S_t + I_t\right) + \begin{bmatrix} b_1 \\ \vdots \\ b_8 \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ \vdots \\ e_{8,t} \end{bmatrix} \tag{2.2}$$

Note that all eight series share the same trend, seasonal and irregular component because the estimate $y_{i,t}$ is obtained from independent subsamples of general population (rotation groups) and therefore, they are just different measures of the same true value of the Australian employment / unemployment level. Trend, seasonal and irregular component (together called "signal") are the components in the model that measure real world changes in the employment / unemployment series. The other two (multivariate) components, rotation group bias and sampling error are related to survey errors that do not reflect real world changes. It is important to separate the sampling error component from the other stochastic components in the model because there is a high chance of identifiability problem. To address this issue, we use the unit-level data from the LFS to estimate parameters of the sampling error (SaE) component and then use them in the state space model (SSM). We describe components of the model together with their state equations in a greater detail below. The Kalman Filter (KF) and

[1] *We denote vectors by bold small letters and matrixes by bold capital letters*

smoother are used for the estimation of latent variables (states) in the model. The model's parameters are estimated by using Maximum Likelihood Estimation (MLE).

a) Trend component T_t is modelled as smooth trend (deterministic trend level T_t and stochastic trend slope R_t) with the following state equation:

$$T_t = T_{t-1} + R_{t-1}, \tag{2.3}$$

$$R_t = R_{t-1} + \zeta_t, \quad \zeta_t \cong NID\left(0, \sigma_\zeta^2\right). \tag{2.4}$$

b) Seasonal component S_t is modelled as stochastic trigonometric form[2]: as:

$$S_t = \sum_{k=1}^{\lfloor s/2 \rfloor} S_{k,t} \ , \tag{2.5}$$

where $S_{k,t}$ is the nonstationary stochastic cycle that has the following state equation:

$$\begin{bmatrix} S_{k,t} \\ S_{k,t}^* \end{bmatrix} = \begin{bmatrix} \cos(2k\pi/s) & \sin(2k\pi/s) \\ -\sin(2k\pi/s) & \cos(2k\pi/s) \end{bmatrix} \begin{bmatrix} S_{k,t-1} \\ S_{k,t-1}^* \end{bmatrix} + \begin{bmatrix} \omega_{k,t} \\ \omega_{k,t}^* \end{bmatrix}, \tag{2.6}$$

with $k = 1, 2, \ldots, \lfloor s/2 \rfloor$ and $\omega_{k,t} \sim NID\left(0, \sigma_\omega^2 \mathbf{I}\right)$ with $\mathrm{E}\left[\omega_{k,t}, \omega_{l,t}\right] = 0$ for $k \neq l$.

c) Irregular component I_t is modelled as white noise $I_t = \xi_t$, $\xi_t \cong NID\left(0, \sigma_\xi^2\right)$.

The initial condition for trend, seasonal and irregular component is diffuse prior (zero initial state value and infinite state variance).

d) Rotation group bias (RGB) effects $b_1, b_2, b_3, b_4, b_5, b_6, b_8$ are permanent level shifts in LFS estimates for waves 1,2,3,4,5,6,8 compared to wave 7 ($b_7 = 0$) respectively. An assumption of deterministic RGB was tested on real data. Firstly, we let RGB for each wave to be time varying (random walk). However, variance of RGB disturbance was almost zero. This supports an assumption of deterministic RGBs. The initial condition for RGB states is also diffuse (zero initial state value and infinite variance). Note we assume that RGB effects are independent so covariances in the initial state variance-covariance matrix are all zeros.

e) Sampling error (SaE) $e_{i,t}$ is modelled as white noise for wave 1 assuming no correlation with previous rotated out rotation group, AR(1) for wave 2 assuming correlation of rotation group that is second month in survey with its previous month estimate and AR(2) for waves 3-8 assuming correlation with previous two months estimates:

$$\begin{aligned} &e_{1,t} = u_{1,t}, \ \ u_{1,t} \cong NID\left(0, \sigma_{u_1}^2\right) \\ &e_{2,t} = \varphi e_{1,t-1} + u_{2,t}, \ \ u_{2,t} \cong NID\left(0, \sigma_{u_2}^2\right) \\ &e_{i,t} = \varphi_1 e_{i-1,t-1} + \varphi_2 e_{i-2,t-2} + u_{i,t}, \ \ u_{i,t} \cong NID\left(0, \sigma_{u_i}^2\right) \end{aligned} \tag{2.7}$$

[2] *Trigonometric form of seasonal was used in SAS. In R we used a dummy variable form and it (almost) did not make any difference in the results*

We assume there is no correlation between wave one (new rotation groups) and previous wave eight (rotating out RGs) although we know that there is some correlation between them due to the fact that both old and new RGs are selected from the same small geographical area. We found that ignoring this correlation does not influence the modelling results much but it makes it simpler to model SaE. Note that SaE within each wave is uncorrelated due to the fact that estimates within each wave belong to different RGs which are independently selected. There is also no correlation across waves at time point t for the same reason. The only SaE correlation is between SaE from the current wave and the previous wave (or two waves) in the previous time point(s). In state space form, SaE is rewritten as:

$$\begin{bmatrix} \mathbf{e_t} \\ \mathbf{e_{t-1}} \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi_1} & \mathbf{\Phi_2} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{e_{t-1}} \\ \mathbf{e_{t-2}} \end{bmatrix} + \begin{bmatrix} \mathbf{u_t} \\ \mathbf{0} \end{bmatrix} \tag{2.8}$$

where $\mathbf{e_t}$ and $\mathbf{u_t}$ are 8-dimentional vectors of SaEs and SaE disturbances respectively and $\mathbf{\Phi_1}$ and $\mathbf{\Phi_2}$ are blocks in SaE transition matrix that have AR coefficients at lower and two level lower diagonals[3]

$$\mathbf{\Phi_1} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ \varphi & 0 & 0 & \dots & 0 \\ 0 & \varphi_1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} = \mathbf{LDiag}\left(\boldsymbol{\varphi_1}\right),\ \mathbf{\Phi}_2 = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \varphi_2 & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} = \mathbf{L^2Diag}\left(\boldsymbol{\varphi_2}\right). \tag{2.9}$$

AR coefficients and SaE variances are estimated from the LFS unit-level data in the following way. Firstly, we estimated autocorrelations (AC) with the method of pseudo-errors suggested by Pfeffermann et. al. (1998) and also with GREG weighted residuals method (see details in Appendix 1). Both methods gave almost identical results. RG autocorrelations of different lags (1-7) are estimated for every month across ten years and then they are combined into wave autocorrelation series. As they do not vary much with time, we average them across time for each combination of wave and lag. From Table 2.1, it can be seen that autocorrelation depends on lag and it does not vary much across waves (for the same lag). Therefore, we averaged them across waves. The corresponding partial autocorrelations (PAC) are only significant at lag 1 and 2. This let us conclude AR(2) process is sufficient for modelling wave level SaEs. The PACs and corresponding AR coefficients are obtained by solving the Yule-Walker equations. Thus, the estimated AR coefficients are $\varphi = 0.589$, $\varphi_1 = 0.466$ and $\varphi_2 = 0.208$ for unemployment and $\varphi = 0.835$, $\varphi_1 = 0.585$ and $\varphi_2 = 0.3$ for employment.

---

[3] ***Diag(x)*** *denotes a diagonal matrix with values from vector* **x** *on the main diagonal;* **LDiag(x)** *denotes a lower diagonal matrix, i.e. with values* **x** *on the entries below the main diagonal, and* **UDiag(x)** *is a corresponding upper diagonal matrix;* $\mathbf{L^2}$**Diag(x)** *denotes a two levels lower diagonal matrix and* $\mathbf{U^2}$**Diag(x)** *is a corresponding upper diagonal matrix.*

2.1 Autocorrelations and partial autocorrelations for wave SEs

a) unemployment

		lag 1	lag 2	lag 3	lag 4	lag 5	lag 6	lag 7
wave 2	AC	0.555						
	PAC	0.555						
wave 3	AC	0.545	0.452					
	PAC	0.545	0.220					
wave 4	AC	0.475	0.443	0.409				
	PAC	0.475	0.281	0.174				
wave 5	AC	0.622	0.442	0.487	0.382			
	PAC	0.622	0.091	0.296	-0.054			
wave 6	AC	0.694	0.505	0.352	0.450	0.393		
	PAC	0.694	0.044	-0.027	0.395	-0.120		
wave 7	AC	0.557	0.535	0.487	0.456	0.399	0.385	
	PAC	0.557	0.326	0.169	0.110	0.025	0.051	
wave 8	AC	0.674	0.519	0.503	0.439	0.468	0.336	0.356
	PAC	0.674	0.118	0.214	0.026	0.203	-0.193	0.210
mean	AC	0.589	0.483	0.447	0.432	0.420	0.361	0.356
	PAC	0.589	0.208	0.160	0.131	0.105	0.005	0.064

b) employment

		lag 1	lag 2	lag 3	lag 4	lag 5	lag 6	lag 7
wave 2	AC	0.820						
	PAC	0.820						
wave 3	AC	0.867	0.798					
	PAC	0.867	0.187					
wave 4	AC	0.861	0.829	0.762				
	PAC	0.861	0.338	-0.008				
wave 5	AC	0.782	0.761	0.773	0.663			
	PAC	0.782	0.383	0.322	-0.133			
wave 6	AC	0.865	0.749	0.722	0.693	0.596		
	PAC	0.865	0.003	0.291	0.020	-0.191		
wave 7	AC	0.852	0.779	0.701	0.693	0.675	0.553	
	PAC	0.852	0.196	0.001	0.229	0.085	-0.391	
wave 8	AC	0.801	0.815	0.740	0.651	0.612	0.568	0.485
	PAC	0.801	0.483	0.059	-0.201	-0.019	0.119	-0.100
mean	AC	0.835	0.788	0.740	0.675	0.627	0.560	0.485
	PAC	0.835	0.300	0.102	-0.035	0.002	-0.074	-0.105

SaE variance is also calculated for each RG in each month for the same period of time by GREG weighted residuals method, which is applied to the unit-level LFS data. SaE variance is relatively consistent across time apart from a small shift during the sample redesign in 2013 and the short period in 2008-2009 when the LFS experienced a significant reduction in sample size. SaE variance (on relative scale) is then combined into waves and averaged across time to form relative standard errors (RSEs) (Table 2.2). As it can be seen, the RSEs are very consistent across waves. Therefore, we assume $\sigma_{e^1}^2 = \sigma_{e^2}^2 = \ldots = \sigma_{e^8}^2 = \sigma_e^2$ where σ_e^2 is an average of estimated RSEs on log scale $\sigma_e^2 = \left(\ln\left(RSE + 1\right)\right)^2$.

2.2 RSEs by wave, %

	1	2	3	4	5	6	7	8	Overall
unemp	6.3	6.5	6.5	6.6	6.6	6.6	6.7	6.7	6.6
emp	0.9	0.9	0.9	1.0	0.9	0.9	0.9	0.9	0.9

SaE variance σ_e^2 is used in SaE disturbance covariance matrix. For wave 1, SaE and SaE disturbance variances are identical as SaE is just white noise in wave 1. For other waves, SaE variance is scaled to take into account SaE autoregressive processes (see equations 2.7-2.8). SaE disturbance variance-covariance matrix is:

$$\mathbf{Q} \triangleq \mathbf{Var}(\mathbf{u_t}) = \begin{bmatrix} \mathbf{Q_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{2.10}$$

$$\mathbf{Q_1} = \begin{bmatrix} \sigma_{u_1}^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_{u_2}^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_{u_3}^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & \sigma_{u_8}^2 \end{bmatrix} = \begin{bmatrix} \sigma_{e_1}^2 & 0 & 0 & \cdots & 0 \\ 0 & \gamma_1\sigma_{e_2}^2 & 0 & \cdots & 0 \\ 0 & 0 & \gamma_2\sigma_{e_3}^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & \gamma_2\sigma_{e_8}^2 \end{bmatrix} = \sigma_e^2\mathbf{Diag}(\boldsymbol{\gamma}) \tag{2.11}$$

where loading factors are denoted as

$$\boldsymbol{\gamma} = \begin{cases} 1, & wave\ 1 \\ \gamma_1, & wave\ 2 \\ \gamma_2, & waves\ 3-8 \end{cases} = \begin{cases} 1, & wave\ 1 \\ (1-\varphi^2), & wave\ 2 \\ \dfrac{1+\varphi_2}{1-\varphi_2}\left[(1-\varphi_2)^2-\varphi_1^2\right], & waves\ 3-8 \end{cases}. \tag{2.12}$$

For the Kalman Filter initialisation of SaE state, it should be taken into account that SaE is a stationary process so the initial condition is exact (non-diffuse) namely zero initial state value with the initial state variance-covariance matrix as the following:

$$\mathbf{Var}\begin{bmatrix} \mathbf{e_t} \\ \mathbf{e_{t-1}} \end{bmatrix} = \mathbf{P_0} = \begin{bmatrix} \mathbf{P_{01}} & \mathbf{P_{02}} \\ \mathbf{P'_{02}} & \mathbf{P_{01}} \end{bmatrix} = \sigma_e^2 \begin{bmatrix} \mathbf{I} & \varphi\mathbf{LDiag}(\mathbf{1}) \\ \varphi\mathbf{UDiag}(\mathbf{1}) & \mathbf{I} \end{bmatrix} \tag{2.13}$$

$$\mathbf{P_{01}} = \begin{bmatrix} \sigma_e^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_e^2 \end{bmatrix} = \sigma_e^2\mathbf{I} \tag{2.14}$$

$$\mathbf{P_{02}} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \varphi\sigma_e^2 & 0 & 0 & \cdots & 0 \\ 0 & \varphi\sigma_e^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} = \varphi\sigma_e^2 \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} = \varphi\sigma_e^2\mathbf{LDiag}(\mathbf{1}) \tag{2.15}$$

where lag 1 autocorrelation φ and SaE variance $\sigma_{e_1}^2 = \ldots = \sigma_{e_8}^2 = \sigma_e^2$ are known. Note that the initial SaE variance-covariance matrix has non-zero variances on the main diagonal and covariances on the lower diagonal to take into account the dependency of states across one-month lagged waves.

Model hyperparameters (σ_ζ^2, σ_ω^2, σ_ξ^2) estimated by MLE are presented in Table 2.3.

2.3 ML estimates of model hyper-parameters (on log scale)

	σ_ζ^2	σ_ω^2	σ_ξ^2
unemp	4.53E-05	7.00E-07	2.60E-04
emp	6.21E-08	1.99E-08	1.90E-06

The residuals diagnostic tests of the above model show that it satisfies the assumption of no serial correlation, homoskedasticity and normality with reasonable trend, seasonal and irregular estimates.

3. STATISTICAL IMPACT MEASUREMENT (SIM) MODELS FOR PARALLEL COLLECTION AND PHASE-IN

3.1 SIM model for parallel collection – difference model

If parallel collection is a possible option for SIM (as it is potentially in the LFS), then there will be two sets of series available during the parallel collection period – from the old and new approaches. We refer to them hereon as control and treatment series respectively. Let us denote the series for control sample as $\mathbf{y_t^C}$ and treatment series as $\mathbf{y_t^T}$. Control series are modelled as

$$\mathbf{y_t^C} = \mathbf{I}_{[8\times 8]}\left(T_t + S_t + I_t\right) + \mathbf{b} + \mathbf{e_t^C} \tag{3.1}$$

and treatment series have a different SaE component and an additional regression component that introduces an impact:

$$\mathbf{y_t^T} = \mathbf{I}_{[8\times 8]}\left(T_t + S_t + I_t\right) + \mathbf{b} + \boldsymbol{\alpha} + \mathbf{e_t^T}. \tag{3.2}$$

Subtracting control series from treatment, we obtain a difference series $\mathbf{y_t^D} = \mathbf{y_t^T} - \mathbf{y_t^C}$ that have only two components – impact $\boldsymbol{\alpha}$ and the difference between treatment and control SaE $\mathbf{e_t^D} = \mathbf{e_t^T} - \mathbf{e_t^C}$:

$$\mathbf{y_t^D} = \boldsymbol{\alpha} + \mathbf{e_t^D} \tag{3.3}$$

or in an expanded form:

$$\begin{bmatrix} y_{1,t}^D \\ \vdots \\ y_{8,t}^D \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_8 \end{bmatrix} + \begin{bmatrix} e_{1,t}^D \\ \vdots \\ e_{8,t}^D \end{bmatrix} \tag{3.4}$$

Note that because treatment and control series share the same trend, seasonal, irregular and RGB, they disappear in the difference series. We refer to the model (3.3) as the difference model (Zhang *et. al.*, 2018). We now consider the components in the model in more detail.

a) Impacts α_i are modelled as permanent wave-specific level shifts.

Initial condition for KF is diffuse prior namely zero initial value and infinite variance. Note that covariances in the initial condition are zeros, i.e. an assumption of independence of impacts across waves is made.

b) State equation for difference in SaE can be written as:

$$\begin{bmatrix} \mathbf{e}_\mathbf{t}^\mathbf{D} \\ \mathbf{e}_\mathbf{t-1}^\mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 \\ \mathbf{1} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{e}_\mathbf{t-1}^\mathbf{D} \\ \mathbf{e}_\mathbf{t-2}^\mathbf{D} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_\mathbf{t}^\mathbf{D} \\ \mathbf{0} \end{bmatrix}. \tag{3.5}$$

where matrix-blocks $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are the same as the real data model (eq. 2.9) and $\mathbf{u}_\mathbf{t}^\mathbf{D}$ is a vector of difference in SaE disturbances ($\mathbf{u}_\mathbf{t}^\mathbf{D} = \mathbf{u}_\mathbf{t}^\mathbf{T} - \mathbf{u}_\mathbf{t}^\mathbf{C}$).

Disturbance variance-covariance matrix for difference in SaE has variances of difference in SaE disturbances on the main diagonal of the upper left matrix-block:

$$\mathbf{Q}^\mathbf{D} = \begin{bmatrix} \mathbf{Q}_1^\mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{3.6}$$

$$\mathbf{Q}_1^\mathbf{D} = \begin{bmatrix} \sigma_{u_1^D}^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_{u_2^D}^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_{u_3^D}^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & \sigma_{u_8^D}^2 \end{bmatrix} = \left(1/\kappa^{PR} + 1 - 2\rho / \sqrt{\kappa^{PR}}\right)\sigma_e^2 \mathbf{Diag}(\boldsymbol{\gamma}) \tag{3.7}$$

where κ^{PR} is a ratio between sample size of treatment and control group during parallel collection.

Note that the variance of difference in SaE disturbance is calculated as

$$\sigma_{u_i^D}^2 = \sigma_{u_i^T}^2 + \sigma_{u_i^C}^2 - 2\rho\sigma_{u_i^T}\sigma_{u_i^C} = \gamma_i\sigma_e^2 / \kappa^{PR} + \gamma_i\sigma_e^2 - 2\rho\gamma_i\sigma_e^2 / \sqrt{\kappa^{PR}} =$$
$$= \left(1/\kappa^{PR} + 1 - 2\rho / \sqrt{\kappa^{PR}}\right)\sigma_e^2\gamma_i$$

where ρ is the correlation between sampling error of treatment and control sample. This correlation is non-zero if treatment and control samples are selected from the same small geographical areas.

KF prior for difference in SaE states has zero initial values and the following initial state variance-covariance matrix:

$$\mathbf{Var}\begin{bmatrix}\mathbf{e}_{\mathbf{t}}^{\mathbf{D}}\\ \mathbf{e}_{\mathbf{t-1}}^{\mathbf{D}}\end{bmatrix}=\mathbf{P}_{\mathbf{0}}^{\mathbf{D}}=\begin{bmatrix}\mathbf{P}_{\mathbf{01}}^{\mathbf{D}} & \mathbf{P}_{\mathbf{02}}^{\mathbf{D}}\\ \mathbf{P}_{\mathbf{02}}^{\prime\mathbf{D}} & \mathbf{P}_{\mathbf{01}}^{\mathbf{D}}\end{bmatrix}= \\ =\left(1/\kappa^{PR}+1-2\rho/\sqrt{\kappa^{PR}}\right)\sigma_e^2\begin{bmatrix}\mathbf{I} & \varphi\mathbf{LDiag}(\mathbf{1})\\ \varphi\mathbf{UDiag}(\mathbf{1}) & \mathbf{I}\end{bmatrix}, \tag{3.8}$$

with variance of difference in SaE on the main diagonal and lagged one covariances across waves. As before, autoregressive coefficients φ, φ_1 and φ_2, the ratio of sample size of treatment and control sample κ^{PR} and SaE variance σ_e^2 are known (see 2.3). However, correlation between SaE of treatment and control sample ρ is unknown and it needs to be estimated.

Therefore, the only parameter that needs to be estimated by MLE in this model is the correlation between sampling error of treatment and control group ρ. Other (hyper-) parameters are known. Note that this is different to Zhang *et. al.* (2018) where difference in SaE disturbance variance $\sigma_{u^D}^2$ rather than ρ is estimated by MLE in the model.

The difference model can be used for estimation of an impact during parallel collection but it is not suitable for cases with no parallel collection or after parallel collection when impact estimates can still be improved by using only treatment series.

3.2 SIM model for phase-in with no parallel collection

As noted above the difference model is not suitable in cases with no parallel collection or after parallel collection because the difference series are defined only in periods where both control and treatment series are available. Therefore different models are needed for those cases.

In the case with no parallel collection, the LFS time series consists of the old (control) series $\mathbf{y}_{\mathbf{t}}^{\mathbf{C}}$ up to the introduction of the new approach and the new (treatment) series $\mathbf{y}_{\mathbf{t}}^{\mathbf{T}}$ continues after that. The model for such combined series can be formulated as:

$$\mathbf{y}_{\mathbf{t}}^{\mathbf{L}}=\mathbf{I}_{[\mathbf{8}\times\mathbf{8}]}\left(T_t+S_t+I_t\right)+\mathbf{b}_{\mathbf{t}}+\mathbf{e}_{\mathbf{t}}^{\mathbf{L}}+\boldsymbol{\alpha}\mathbf{x}_{\mathbf{t}} \tag{3.9}$$

or in an expanded form:

$$\begin{bmatrix}y_{1,t}^L\\ \vdots\\ y_{8,t}^L\end{bmatrix}=\mathbf{I}_{[\mathbf{8}\times\mathbf{8}]}\left(T_t+S_t+I_t\right)+\begin{bmatrix}b_1\\ \vdots\\ b_8\end{bmatrix}+\begin{bmatrix}e_{1,t}^L\\ \vdots\\ e_{8,t}^L\end{bmatrix}+\begin{bmatrix}\alpha_1x_{1,t}\\ \vdots\\ \alpha_8x_{8,t}\end{bmatrix}, \tag{3.10}$$

where:

$$y_{i,t}^L=\begin{cases}y_{i,t}^C & \text{before wave } i \text{ phase-in}\\ y_{i,t}^T & \text{since wave } i \text{ phase-in}\end{cases},$$

$$x_{i,t} = \begin{cases} 0 & \text{before wave } i \text{ phase-in} \\ 1 & \text{since wave } i \text{ phase-in} \end{cases}.$$

Note that dummy variables $x_{i,t}$ are different for all waves because it is expected that the new approach will be introduced incrementally by one rotation group a month. This process is called phase-in (see 5.1 for details).

Trend, seasonal, irregular component and RGB are modelled as described in 2.3 (real data model) and the impact component is the same as the difference model (see 3.1). We describe the SaE component in more detail below.

a) SaE state equation is:

$$\begin{bmatrix} \mathbf{e}_\mathbf{t}^\mathbf{L} \\ \mathbf{e}_\mathbf{t-1}^\mathbf{L} \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_\mathbf{1} & \mathbf{\Phi}_\mathbf{2} \\ \mathbf{1} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{e}_\mathbf{t-1}^\mathbf{L} \\ \mathbf{e}_\mathbf{t-2}^\mathbf{L} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_\mathbf{t}^\mathbf{L} \\ \mathbf{0} \end{bmatrix}. \tag{3.11}$$

Disturbance variance-covariance matrix for SaE has variance of SaE disturbance on the main diagonal of the upper left matrix-block:

$$\mathbf{Q}^\mathbf{L} = \begin{bmatrix} \mathbf{Q}_\mathbf{1}^\mathbf{L} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{3.12}$$

$$\mathbf{Q}_\mathbf{1}^\mathbf{L} = \begin{bmatrix} \sigma_{e_1}^2 / \kappa_{1,t} & 0 & 0 & \cdots & 0 \\ 0 & \gamma_1 \sigma_{e_2}^2 / \kappa_{2,t} & 0 & \cdots & 0 \\ 0 & 0 & \gamma_2 \sigma_{e_3}^2 / \kappa_{3,t} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & \gamma_2 \sigma_{e_8}^2 / \kappa_{8,t} \end{bmatrix} = \sigma_e^2 \mathbf{Diag}(\boldsymbol{\gamma} / \boldsymbol{\kappa}) \tag{3.13}$$

where $\boldsymbol{\kappa}$ is a vector of dummy variables $\kappa_{i,t}$ specified as:

$$\kappa_{i,t} = \begin{cases} \kappa^{PI} & \text{if wave } i \text{ has treatment group (phase-in) at time } t \\ 1 & \text{otherwise} \end{cases}$$

and κ^{PI} is the ratio between sample size of treatment and control group during phase-in.

Initial condition for KF is the same as real data model, namely zero initial state value and initial state variance-covariance matrix (see eq. 2.13):

$$\mathbf{Var}\begin{bmatrix} \mathbf{e}_\mathbf{t}^\mathbf{L} \\ \mathbf{e}_\mathbf{t-1}^\mathbf{L} \end{bmatrix} = \mathbf{P}_\mathbf{0}^\mathbf{L} = \mathbf{P}_\mathbf{0} = \sigma_e^2 \begin{bmatrix} \mathbf{I} & \varphi \mathbf{LDiag}(\mathbf{1}) \\ \varphi \mathbf{UDiag}(\mathbf{1}) & \mathbf{I} \end{bmatrix}. \tag{3.14}$$

The hyperparameters estimated by MLE are the same as the ones estimated in the model for real data (i.e. σ_ζ^2, σ_ω^2, σ_ξ^2).

3.3 SIM model for phase-in with parallel collection

In the case with parallel collection is available, estimates of impacts can still be improved after the phase-in period commences even without having the control series. The model for impact measurement in this case is exactly the same as the phase-in model (see eq. 3.9). However, information about the size of impacts obtained from the parallel collection can be used for the KF initialisation of the impact states.

Unlike the difference and phase-in model, an initial condition of the impacts is non-diffuse. The difference model estimates of the impacts and the respective variance-covariance matrix are taken as the initial condition for KF. Note that covariance of the impacts is not necessarily zero as it is for the difference model.

4. STATISTICAL IMPACT MEASUREMENT WITH THE FULL MODEL

4.1 Full model for SIM

The obvious disadvantage of the difference model is that it is only applicable when there is a parallel collection and a different model has to be used after that. However, the full model overcomes this disadvantage as it uses both control and parallel collection information. It performs significantly better than the difference model in terms of impact detection as it will be shown in section 5. The following shows the model specification of the full model:

$$\begin{bmatrix} \mathbf{y}_\mathbf{t}^\mathbf{C} \\ \mathbf{y}_\mathbf{t}^\mathbf{T} \end{bmatrix} = \mathbf{I}_{[16\times16]}\left(T_t + S_t + I_t\right) + \mathbf{b} + \begin{bmatrix} \mathbf{e}_\mathbf{t}^\mathbf{C} \\ \mathbf{e}_\mathbf{t}^\mathbf{T} + \boldsymbol{\alpha} \end{bmatrix} \quad (4.1)$$

or in an expanded form:

$$\begin{bmatrix} y_{1,t}^C \\ \vdots \\ y_{8,t}^C \\ y_{1,t}^T \\ \vdots \\ y_{8,t}^T \end{bmatrix} = \mathbf{I}_{[16\times16]}\left(T_t + S_t + I_t\right) + \begin{bmatrix} b_1 \\ \vdots \\ b_8 \\ b_1 \\ \vdots \\ b_8 \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \alpha_1 \\ \vdots \\ \alpha_8 \end{bmatrix} + \begin{bmatrix} e_{1,t}^C \\ \vdots \\ e_{8,t}^C \\ e_{1,t}^T \\ \vdots \\ e_{8,t}^T \end{bmatrix}. \quad (4.2)$$

Note that treatment series have missing values before parallel collection (or before phase-in in the case of no parallel collection) and control series have missing values since phase-in commences.

The trend, seasonal, irregular and RGB component are modelled the same as 2.3 and the impact component is modelled in the same way as the difference model (see 3.1). We describe the SaE component, which is different from the models described above, of the full model as follows:

a) SaE state equation is:

$$\begin{bmatrix} \mathbf{e}_t^C \\ \mathbf{e}_t^T \\ \mathbf{e}_{t-1}^C \\ \mathbf{e}_{t-1}^T \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi}_1 & \mathbf{0} & \boldsymbol{\Phi}_2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}_1 & \mathbf{0} & \boldsymbol{\Phi}_2 \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{e}_{t-1}^C \\ \mathbf{e}_{t-1}^T \\ \mathbf{e}_{t-2}^C \\ \mathbf{e}_{t-2}^T \end{bmatrix} + \begin{bmatrix} \mathbf{u}_t^C \\ \mathbf{u}_t^T \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \tag{4.3}$$

The variance of SaE disturbance for the control and treatment group is on the main diagonal in the variance-covariance matrix of the disturbance. Note that the variance of SaE disturbance for treatment group needs to be scaled to take into account the fact that the sample size of treatment group (particularly during parallel collection period) can be smaller than the sample size of control sample. Note that $\mathbf{Q}_1^F = \mathbf{Q}_1$.

$$\mathbf{Q}^F = \begin{bmatrix} \mathbf{Q}_1^F & \mathbf{Q}_3^F & \mathbf{0} & \mathbf{0} \\ \mathbf{Q}_3^F & \mathbf{Q}_2^F & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} = \sigma_e^2 \begin{bmatrix} \mathbf{Diag}(\boldsymbol{\gamma}) & \rho\mathbf{Diag}\left(\boldsymbol{\gamma}/\sqrt{\boldsymbol{\kappa}^*}\right) & \mathbf{0} & \mathbf{0} \\ \rho\mathbf{Diag}\left(\boldsymbol{\gamma}/\sqrt{\boldsymbol{\kappa}^*}\right) & \mathbf{Diag}(\boldsymbol{\gamma}/\boldsymbol{\kappa}^*) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{4.4}$$

where $\boldsymbol{\kappa}^*$ is a vector of variables $\kappa_{i,t}^*$ specified as:

$$\kappa_{i,t}^* = \begin{cases} \kappa^{PR} \text{ if wave } i \text{ has treatment group (parallel run) at time } t \\ \kappa^{PI} \quad \text{if wave } i \text{ has treatment group (phase-in) at time } t \\ 1 \qquad\qquad \text{otherwise} \end{cases},$$

KF initial condition for SaE states is the following: zero initial state values with the initial state variance-covariance matrix

$$\mathbf{Var}\begin{bmatrix} \mathbf{e}_t^C \\ \mathbf{e}_t^T \\ \mathbf{e}_{t-1}^C \\ \mathbf{e}_{t-1}^T \end{bmatrix} = \mathbf{P}_0^F = \sigma_e^2 \begin{bmatrix} \mathbf{I} & \left(\rho/\sqrt{\kappa^{PR}}\right)\mathbf{I} & \varphi\mathbf{LDiag}(\mathbf{1}) & \left(\rho\varphi/\sqrt{\kappa^{PR}}\right)\mathbf{LDiag}(\mathbf{1}) \\ \left(\rho/\sqrt{\kappa^{PR}}\right)\mathbf{I} & \left(1/\kappa^{PR}\right)\mathbf{I} & \left(\rho\varphi/\sqrt{\kappa^{PR}}\right)\mathbf{LDiag}(\mathbf{1}) & \left(\varphi/\kappa^{PR}\right)\mathbf{LDiag}(\mathbf{1}) \\ \varphi\mathbf{UDiag}(\mathbf{1}) & \left(\rho\varphi/\sqrt{\kappa^{PR}}\right)\mathbf{UDiag}(\mathbf{1}) & \mathbf{I} & \left(\rho/\sqrt{\kappa^{PR}}\right)\mathbf{I} \\ \left(\rho\varphi/\sqrt{\kappa^{PR}}\right)\mathbf{UDiag}(\mathbf{1}) & \left(\varphi/\kappa^{PR}\right)\mathbf{UDiag}(\mathbf{1}) & \left(\rho/\sqrt{\kappa^{PR}}\right)\mathbf{I} & \left(1/\kappa^{PR}\right)\mathbf{I} \end{bmatrix} \tag{4.5}$$

As before all parameters apart from ρ for SaE are known.

The estimated parameters in this model are: σ_ζ^2, σ_ω^2, σ_ξ^2 and ρ.

4.2 Alternative formulation of the full model

Alternatively the full model can be formulated as:

$$\begin{aligned} \mathbf{y}_t^C &= \mathbf{I}_{[8\times 8]}\left(T_t + S_t + I_t\right) + \mathbf{b} + \mathbf{e}_t^C \\ \mathbf{y}_t^D &= \mathbf{e}_t^D + \boldsymbol{\alpha} \end{aligned} \tag{4.6}$$

In this model, the treatment series $\mathbf{y}_t^T$ is replaced by the difference series $\mathbf{y}_t^D$. Unlike the full model, the trend, seasonal, irregular and RGB component are estimated from only the control series. Similar to the difference model, this model can only be applied to estimate impacts during parallel collection period because the difference series are unspecified otherwise.

This form of the full model demonstrates the main difference between the difference and the full model. As it is described, the full model uses all information available from historical time series and parallel collection period, although there are more parameters that need to be estimated than the difference model.

The trend, seasonal, irregular and RGB component are modelled in the same fashion as 2.3 and the impact component is modelled in the same way as the difference model (see 3.1). We describe the SaE component as follows.

a) SaE state equation:

$$\begin{bmatrix} \mathbf{e}_t^C \\ \mathbf{e}_t^D \\ \mathbf{e}_{t-1}^C \\ \mathbf{e}_{t-1}^D \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_1 & \mathbf{0} & \mathbf{\Phi}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi}_1 & \mathbf{0} & \mathbf{\Phi}_2 \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{e}_{t-1}^C \\ \mathbf{e}_{t-1}^D \\ \mathbf{e}_{t-2}^C \\ \mathbf{e}_{t-2}^D \end{bmatrix} + \begin{bmatrix} \mathbf{u}_t^C \\ \mathbf{u}_t^D \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \tag{4.7}$$

Disturbance variance-covariance matrix:

$$\mathbf{Q}^{F*} = \sigma_e^2 \begin{bmatrix} \mathbf{Diag}(\boldsymbol{\gamma}) & \mathbf{Diag}\left(\boldsymbol{\gamma}\left(\rho/\sqrt{\boldsymbol{\kappa}^*}-1\right)\right) & \mathbf{0} & \mathbf{0} \\ \mathbf{Diag}\left(\boldsymbol{\gamma}\left(\rho/\sqrt{\boldsymbol{\kappa}^*}-1\right)\right) & \mathbf{Diag}\left(\boldsymbol{\gamma}\left(1/\boldsymbol{\kappa}^*+1-2\rho/\sqrt{\boldsymbol{\kappa}^*}\right)\right) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{4.8}$$

where covariances are calculated as below:

$$\mathrm{cov}\left(u_{i,t}^C, u_{i,t}^D\right) = \mathrm{cov}\left(u_{i,t}^C, u_{i,t}^T - u_{i,t}^C\right) = \mathrm{cov}\left(u_{i,t}^C, u_{i,t}^T\right) - \mathrm{var}\left(u_{i,t}^C\right) = \rho\gamma_i\sigma_e^2/\sqrt{\kappa_{i,t}^*} - \gamma_i\sigma_e^2 =$$
$$= \sigma_e^2\gamma_i\left(\rho/\sqrt{\kappa_{i,t}^*}-1\right).$$

KF initial condition is zero initial state values with initial state variance-covariance matrix:

$$\mathbf{Var}\begin{bmatrix} \mathbf{e}_t^C \\ \mathbf{e}_t^D \\ \mathbf{e}_{t-1}^C \\ \mathbf{e}_{t-1}^D \end{bmatrix} = \mathbf{P}_0^{F*} = \sigma_e^2 \begin{bmatrix} \mathbf{I} & \mathbf{Diag}\left(\rho/\sqrt{\kappa^{PR}}-1\right) & \varphi\mathbf{LDiag}(1) & \varphi\mathbf{LDiag}\left(\rho/\sqrt{\kappa^{PR}}-1\right) \\ \mathbf{Diag}\left(\rho/\sqrt{\kappa^{PR}}-1\right) & \mathbf{Diag}\left(1/\kappa^{PR}+1-2\rho/\sqrt{\kappa^{PR}}\right) & \varphi\mathbf{LDiag}\left(\rho/\sqrt{\kappa^{PR}}-1\right) & \varphi\mathbf{LDiag}\left(1/\kappa^{PR}+1-2\rho/\sqrt{\kappa^{PR}}\right) \\ \varphi\mathbf{UDiag}(1) & \varphi\mathbf{UDiag}\left(\rho/\sqrt{\kappa^{PR}}-1\right) & \mathbf{I} & \mathbf{Diag}\left(\rho/\sqrt{\kappa^{PR}}-1\right) \\ \varphi\mathbf{UDiag}\left(\rho/\sqrt{\kappa^{PR}}-1\right) & \varphi\mathbf{UDiag}\left(1/\kappa^{PR}+1-2\rho/\sqrt{\kappa^{PR}}\right) & \mathbf{Diag}\left(\rho/\sqrt{\kappa^{PR}}-1\right) & \mathbf{Diag}\left(1/\kappa^{PR}+1-2\rho/\sqrt{\kappa^{PR}}\right) \end{bmatrix}. \tag{4.9}$$

The estimated parameters (MLE) for this model are: σ_ζ^2, σ_ω^2, σ_ξ^2, and ρ.

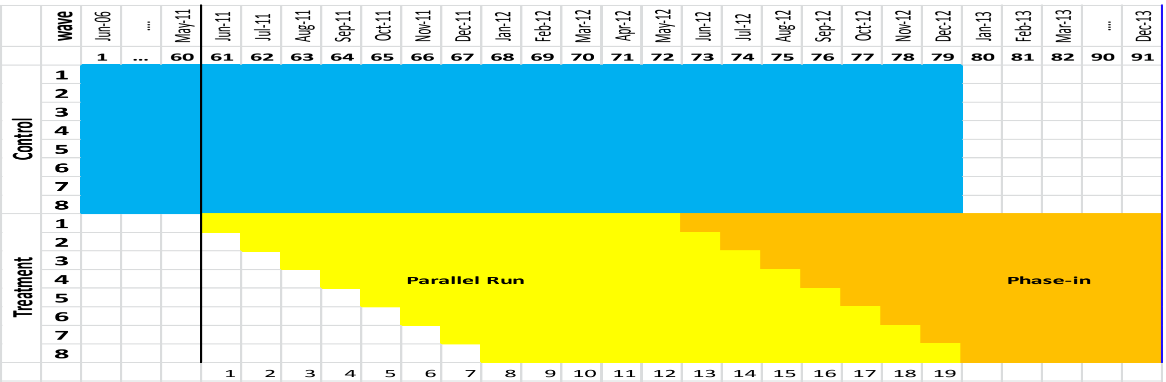
5. DATA SIMULATION AND OVERALL IMPACT ASSESSMENT

5.1 Data simulation

Each series used in this study is simulated with SAS. A hundred pairs of control and treatment series for employment and unemployment are simulated for each rotation group under different correlation between treatment and control samples (the correlation is unknown in real life). RG series are then reformulated into wave series by time in survey. When the ratio between sample size of treatment and control series is different for parallel collection and phase-in periods, we simulate two sets of series with different parallel collection and phase-in ratios. We then replace the treatment series from the first set with the treatment series from the second set when phase-in begins. The simulated data are generated from June 2006 to December 2013. It is expected that the beginning month for actual SIM will be June 2006, however, the last month in series will be the last month before transitioning LFS to new environment. Control series are replaced by missing values from October 2012 and treatment series are replaced by missing values until parallel collection begins, which is in June 2011.

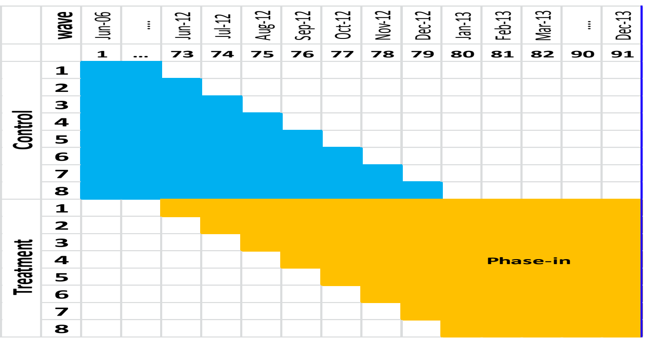
We use the parallel collection scheme that will most likely be implemented in LFS transformation period if parallel collection is available. However, simulated impacts are simulated from June 2011 unlike in real life (the exact time has yet to be decided). There are two schemes under consideration – with 16 and 19 months of parallel collection. The 19 month scheme is presented in Figure 5.1.

5.1 Simulation scheme, 19 months of parallel collection



The scheme without parallel collection is presented in Figure 5.2.

5.2 Simulation scheme in case of no parallel collection



Each simulated series contains a few components that are obtained separately:

1. "Real world component" that is obtained from the national (published) unemployment / employment series (on log scale) by removing the sampling error with the univariate state space model. The same component is used to generate all waves in both treatment and control series.

2. SaE is generated for eight waves as white noise for wave 1, AR(1) process for wave 2 and AR(2) process for waves 3-8 (see Appendix 2 for more details). Note that observations are independent among waves at a particular time point and within waves. Therefore, waves follow an AR process only with their lagged values from the previous waves. AR parameters and variance of SaE disturbance used in the simulation are estimated from real LFS data at the rotation group level (see section 2.3) and they are in log scale. Additionally, treatment series SaE disturbance variance is scaled by a factor that is an inverse of the ratio between sample size of treatment and control group. There are a few scenarios with different ratio for parallel collection and phase-in periods.

3. Rotation group bias (RGB) is estimated from real LFS data as the ratio of GREG estimate for wave i to GREG estimate for wave seven and is added to the series for all waves apart from wave seven (see Table 5.1). Note that in log scale, RGB is a difference rather than a ratio. RGB is estimated across the same ten year period as the other predefined components described above. Monthly GREG estimates are obtained by calibrating each rotation group to the national level benchmarks and then they are averaged across time. Note that wave seven is a reference wave although any wave can be the reference wave or average across waves. The choice of a reference wave does not influence the accuracy of any time series component and the size of the components except the trend level.

4. A wave specific permanent level shift (impact, RGBD) in log scale is added to the treatment series (see Table 5.1). This impact is generated so that the overall impact is equal to one standard error of the published estimate (2.5% for unemployment and 0.37% for employment). Note that the overall impact is obtained by averaging the product of the exponential of wave impacts and the composite estimator weights of the current month CW0 across waves.

5.1 Simulated RGB and impact
a) unemployment

wave	RGB	RGBD (impact)	CW0	exp(RGBD)	CW0*exp(RGBD)
1	0.0814	0.1000	0.7941	1.1052	0.8776
2	0.0499	0.0500	0.9823	1.0513	1.0326
3	0.0376	0.0350	1.0240	1.0356	1.0605
4	0.0260	0.0250	1.0342	1.0253	1.0603
5	0.0059	0.0100	1.0375	1.0101	1.0479
6	0.0156	0.0035	1.0393	1.0035	1.0429
7	0	0.0000	1.0444	1.0000	1.0444
8	-0.0125	-0.0100	1.0444	0.9900	1.0340
				total impact	**2.5%**

b) employment

wave	RGB	RGBD (impact)	CW0	exp(RGBD)	CW0*exp(RGBD)
1	0.0046	0.0500	0.7941	1.0513	0.8348
2	0.0000	0.0250	0.9823	1.0253	1.0071
3	-0.0012	0.0100	1.0240	1.0101	1.0343
4	-0.0008	0.0050	1.0342	1.0050	1.0393
5	0.0009	0.0005	1.0375	1.0005	1.0380
6	0.0002	0.0001	1.0393	1.0001	1.0394
7	0	0	1.0444	1.0000	1.0444
8	0.0001	-0.0510	1.0444	0.9503	0.9924
				total impact	**0.37%**

The data generation procedure described above is repeated 100 times. This gives us 100 replicates of eight control and treatment wave series (8*2*100 series) for each scenario. All series are generated in log scale. Note that all components used in the simulation are obtained or estimated from the LFS data to ensure the simulated series are as “realistic” as possible.

5.2 Overall impact assessment

The main question to answer with our study is the size of the impact that can be detected by the model at the end of parallel collection and at the beginning of phase-in period. The size of detectable impact can be measured by minimum detectable impact ratio (MDIR), which is defined as the size of the impact that can be detected based on a stated accuracy criterion:

$$MDIR = 2.8\frac{SE(\alpha)}{\Delta} \qquad (5.1)$$

where 2.8 is a multiplier derived from a set of predefined type I and II errors (5% and 20% respectively), $SE(\alpha)$ is the standard error (SE) of detected impact obtained from the model and Δ is one SE of unemployment or employment according to ABS LFS publications (on average 2.5% for unemployment and 0.37% for employment).

Alternatively we could fix a size of an impact that we would like to detect (MDIR) and assess the power of detection of that size of the impact. The power then can be estimated as:

$$power = 1-\Phi\left(z_{0.975}-\frac{MDIR\cdot\Delta}{SE(\alpha)}\right)+\Phi\left(-z_{0.975}-\frac{MDIR\cdot\Delta}{SE(\alpha)}\right) \qquad (5.2)$$

where Φ is the standard Normal distribution function.

We can obtain actual empirical impact SEs as well as modelled ones from the simulation results. The purpose of comparing modelled SEs with empirical ones (they are available only in simulation study) is to examine if the modelled SE is an unbiased estimate of the true SE.

Empirical standard errors $s^{\alpha,e}$ of an overall impact are calculated as the standard deviation of the overall impact across 100 replicates:

$$s^{\alpha,e} = \sqrt{\frac{\sum_{j=1}^{100}\left(\alpha_j - \bar{\alpha}\right)^2}{99}} \tag{5.3}$$

where α_j is the overall impact in replicate *j* calculated using the composite estimator weight (BLUE multiplier) for current month w_i and taking exponential of wave level impacts α_{ij}:

$$\alpha_j = \left(\frac{\sum_{i=1}^{8} w_i e^{\alpha_{ij}}}{8} - 1\right) * 100, \tag{5.4}$$

and $\bar{\alpha}$ is the average of α_j over100 replicates. Note that wave level and overall impacts are on different scales: wave impacts are in log scale whereas the overall impact is measured in %.

Modelled SE of an overall impact is estimated as the average SE of an estimated overall impact across 100 replicates:

$$s^{\alpha,m} = \sqrt{\frac{\sum_{j=1}^{100} s_j^{2\ \alpha,m}}{100}} \tag{5.5}$$

where an overall impact SE for replicate *j* is calculated by using this formula:

$$s_j^{\alpha,m} = \sqrt{\operatorname{var}\left(\left(\frac{\sum_{i=1}^{8} w_i e^{\alpha_{ij}}}{8} - 1\right) \cdot 100\right)} = \frac{100c}{8}\sqrt{\sum_{i=1}^{8} w_i^2 \operatorname{var}\left(e^{\alpha_{ij}}\right)}$$

/ Taylor series expansion /

$$\approx \frac{100c}{8}\sqrt{\sum_{i=1}^{8} w_i^2 \left(e^{\alpha_{ij}}\right)^2 \operatorname{var}\left(\alpha_{ij}\right)} = \frac{100c}{8}\sqrt{\sum_{i=1}^{8} w_i^2 \left(e^{\alpha_{ij}}\right)^2 s_{ij}^2}\ , \tag{5.6}$$

where c is a correction factor that accounts for dependency of impacts in waves ($c = 1.92$ for unemployment and 2.46 for employment),

s_{ij} is the SE of an estimated impact in wave *i* of replicate *j* obtained from the model.

General form of correction factor c

If we calculate the variance correctly then

$$\operatorname{var}\left(\sum_{i=1}^{8} w_i e^{\alpha_{ij}}\right) = \sum_{i=1}^{8} w_i^2 \operatorname{var}\left(e^{\alpha_{ij}}\right) + 2\sum_{i=1}^{7}\sum_{k>i}^{8} w_i w_k \operatorname{cov}\left(e^{\alpha_{ij}}, e^{\alpha_{kj}}\right) \approx$$

$$\approx \sum_{i=1}^{8} w_i^2 \left(e^{\alpha_{ij}}\right)^2 \operatorname{var}\left(\alpha_{ij}\right) + 2\sum_{i=1}^{7}\sum_{k>i}^{8} w_i w_k e^{\alpha_{ij}} e^{\alpha_{kj}} \varphi_{k-i} \sqrt{\operatorname{var}\left(\alpha_{ij}\right)\operatorname{var}\left(\alpha_{kj}\right)} \quad (5.7)$$

Ignoring the dependence of wave impact estimates we are getting:

$$\operatorname{var}\left(\sum_{i=1}^{8} w_i e^{\alpha_{ij}}\right) \approx \sum_{i=1}^{8} w_i^2 \left(e^{\alpha_{ij}}\right)^2 \operatorname{var}\left(e^{\alpha_{ij}}\right) \quad (5.8)$$

Then dependence correction factor is defined as:

$$c = \sqrt{\frac{\sum_{i=1}^{8} w_i^2 \left(e^{\alpha_{ij}}\right)^2 \operatorname{var}\left(\alpha_{ij}\right) + 2\sum_{i=1}^{7}\sum_{k>i}^{8} w_i w_k e^{\alpha_{ij}} e^{\alpha_{kj}} \varphi_{k-i} \sqrt{\operatorname{var}\left(\alpha_{ij}\right)\operatorname{var}\left(\alpha_{kj}\right)}}{\sum_{i=1}^{8} w_i^2 \left(e^{\alpha_{ij}}\right)^2 \operatorname{var}\left(e^{\alpha_{ij}}\right)}} \quad (5.9)$$

To make it easier, assume $\operatorname{var}\left(\alpha_{ij}\right) = \sigma_j^2$ across waves and $e^{\alpha_{ij}}$ is approximately 1. Then we get that

$$c = \sqrt{\frac{\sum_{i=1}^{8} w_i^2 + 2\sum_{i=1}^{7}\sum_{k>i}^{8} w_i w_k \varphi_{k-i}}{\sum_{i=1}^{8} w_i^2}} \quad (5.10)$$

for lagged correlations φ_{k-i} defined by the corresponding autoregressive process in sampling errors.

6. COMPARISON OF DIFFERENT SIM MODELS

6.1 Models for parallel collection

There are two models that are tested for measuring impacts during parallel collection – the full model and difference model. Impact estimates and their SEs (empirical and modelled) at the wave level are presented in Appendix 3. There following provides a summary of the results:

- Impact estimates are unbiased (converging to the true value) however, there is a small gap between estimated and true (simulated) impacts with a short parallel collection period.
- The modelled SEs are close to empirical SEs. However, both the full and difference model slightly overestimated the impact SEs.
- The greater the correlation ρ between SaE of treatment and control sample, the smaller the gap between the estimated and simulated impacts. In addition, the impact estimate SEs are lower with a higher correlation.

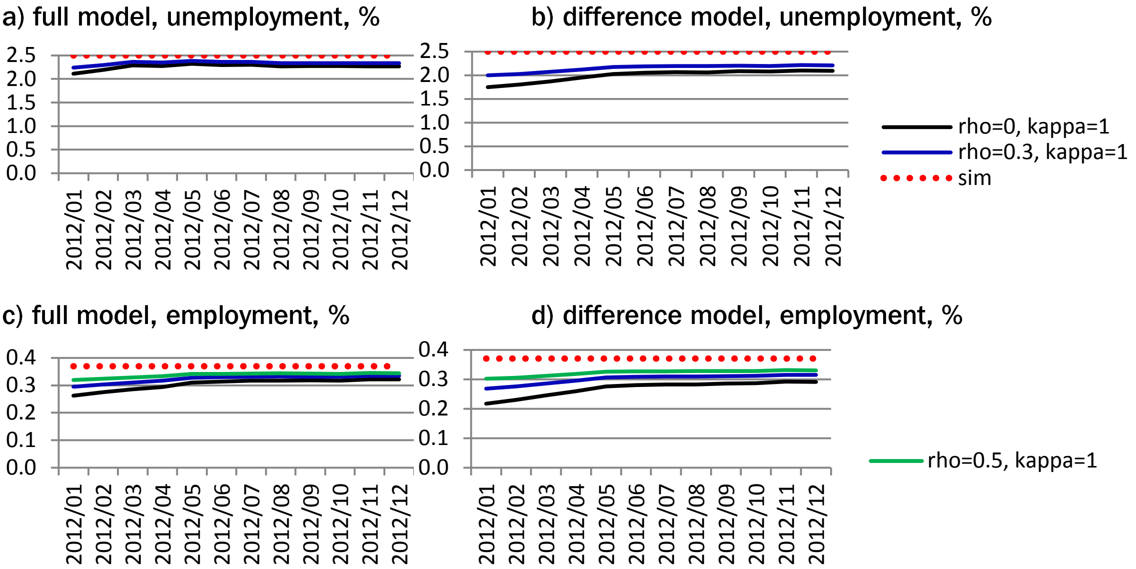
- The smaller the ratio κ^{PR} between sample size of treatment and control group during parallel collection period[4], the greater the gap between the estimated and simulated impacts. In addition, the impact SEs are higher with a lower ratio.
- SEs increase by waves. This is because there is a different number of observations in each wave due to the gradual introduction of the parallel sample (see Figure 5.1). For example, in the eighth month since the beginning of the parallel collection (the first month when all waves have been introduced to the new approach), treatment sample has eight observations for wave 1 and only one observation for wave 8. The difference becomes smaller over time due to a smaller relative difference in the treatment series length of each wave.
- The distribution of correlation estimates ρ (not presented here) across 100 replicates does not violate the assumption of normality significantly. The standard error of correlation estimate depends on the correlation value, i.e. the greater the simulated correlation, the more precise the estimate of correlation is. On average, the full model estimates correlation slightly better than the difference model.
- The full model also outperforms the difference model in terms of the efficiency and the true value convergence rate of the impact estimates. The improvement in the full model performance depends on the length of parallel collection and the correlation between control and treatment SaE (ρ). On average, there is a lesser improvement as the length of parallel collection and correlation increases. E.g. for $\rho = 0$ (unemployment) there is, on average, a 25% and 17% improvement in SE at month 8 and 19 respectively. With employment data, the improvement is slightly smaller (about 20% for rho=0 with 8 months of parallel collection) but the pattern is still the same.
- Alternative formulation of the full model (results are not presented here) performs (almost) the same as the full model.

Overall impact estimates and their standard errors (empirical and modelled) are presented in Figure 6.1-6.2. It can be seen that the conclusions made from the wave level results are also consistent with the overall level results. Note that the modelled SEs are corrected on wave impact estimates dependence. They are close to empirical SEs. Uncorrected SEs (not presented here) are on average twice smaller than corrected ones for unemployment and about 2.5 times smaller for employment.

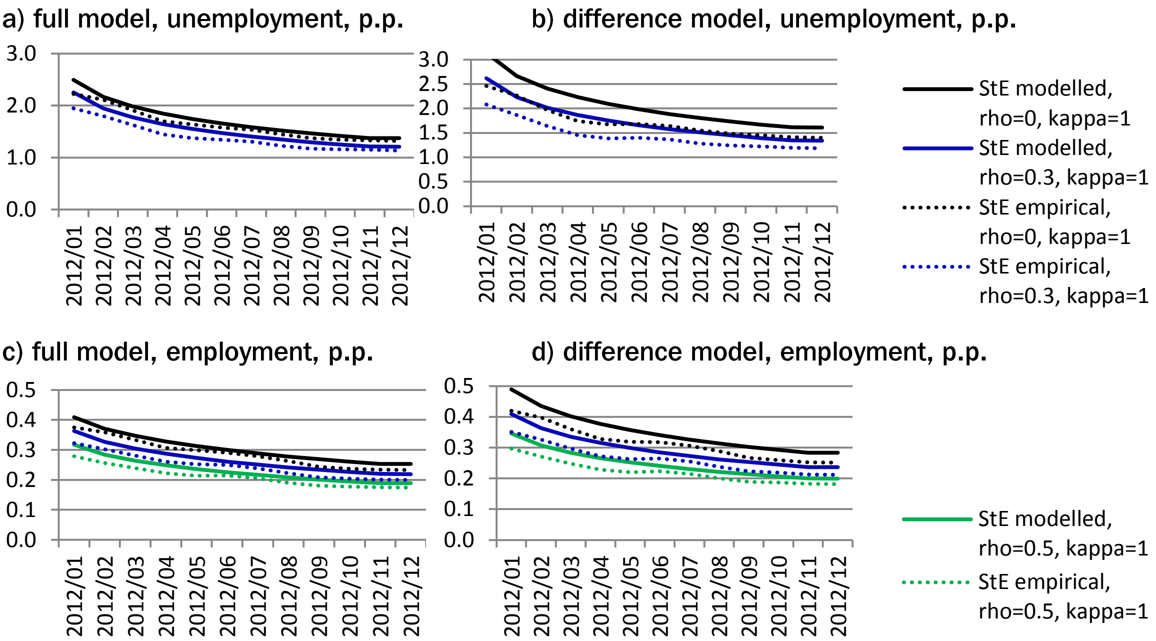
---

[4] *We tested scenarios with* κ^{PR} *=0.3, 0.5, 0.8 and 1 although in Appendix 3, the only result that is presented is the scenario with* κ^{PR} *=1.*

6.1 Simulated (dotted line) and estimated (solid lines) overall impact during parallel collection period[5]



6.2 Empirical (dotted line) and modelled (solid line) overall impact SE during parallel collection period



Empirical power to detect an impact (Table 6.1) reflects the power to detect an impact if the true value of the SE of impact is used. However, we do not have empirical standard errors in practice and therefore, we only have the estimated impact SEs from the model. Modelled SEs (Table 6.2) are slightly overestimated so therefore the power to detect an impact is slightly smaller than the power calculated with empirical (true) SEs. From the full model, the power to detect an impact that is the size of one SE (1 SE = 2.5% for unemployment and 0.37% for employment) is between 45% ($\rho=0$) and 55% ($\rho=0.3$)

[5] *The beginning month at the plots is the eighth month since beginning of parallel collection period (when all eight waves have been phased-in).*

for unemployment depending on correlation ρ (it is expected that correlation will be between 0% and 30%) and about 50% for employment (expected correlation is about 50%). The difference model power is significantly smaller, 34-46% for unemployment and about 46% for employment.

6.1 Empirical power (%) of detecting overall impact in 19 month of parallel collection

		unemployment		employment		
	rho	0	0.3	0	0.3	0.5
	kappa	1	1	1	1	1
1 SE	Difference model	43.1	56.2	31.2	41.4	53.2
	Full model	47.6	59.8	35.6	45.7	56.7
1.5 SE	Difference model	76.5	88.8	59.7	74.4	86.4
	Full model	81.3	91.2	66.5	79.3	89.1
2 SE	Difference model	94.7	98.9	83.7	93.7	98.3
	Full model	96.7	99.3	88.9	95.9	98.9

6.2 Modelled power (%) of detecting overall impact in 19 month of parallel collection

		unemployment		employment		
	rho	0	0.3	0	0.3	0.5
	kappa	1	1	1	1	1
1 SE	Difference model	34.3	46.1	25.7	34.6	45.7
	Full model	44.6	54.5	30.9	39.2	50.0
1.5 SE	Difference model	64.6	79.7	50.0	65.1	79.3
	Full model	78.1	87.5	59.1	71.6	83.6
2 SE	Difference model	87.5	96.1	74.3	87.9	95.9
	Full model	95.4	98.6	83.2	92.1	97.5

6.2 Models for phase-in

1) Case of no parallel collection

In case there is no parallel collection two models can be used for impact measurement – phase-in model and full model. Note that we use phase-in scheme presented in Figure 5.2[6]. The two models give (almost) identical results so below (Figure 6.3-6.6) we present only results from the full model.

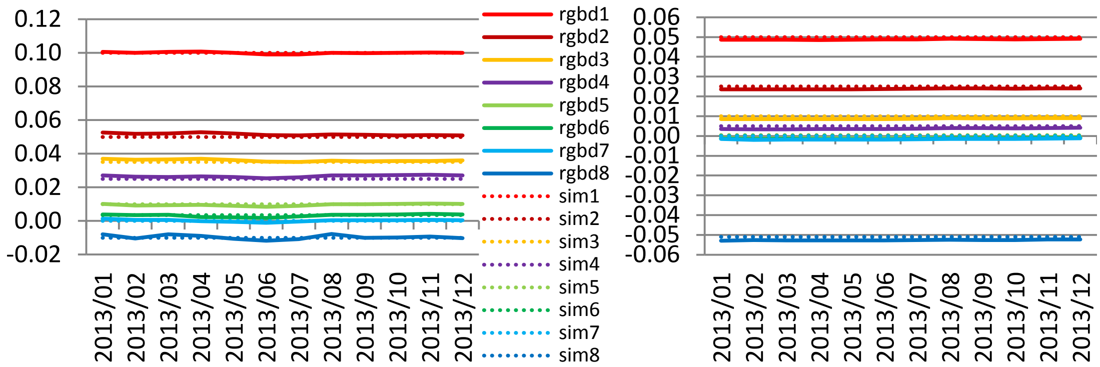
As expected, impact SEs are much greater in case of no parallel collection compared to case with parallel collection at both wave and overall levels. Therefore the power of impact detection is much greater for the case with parallel collection (Table 6.3). Such in case of no correlation between treatment and control SaEs the power to detect an impact of size of one SE of unemployment is about 18% without parallel collection (19 months since phase-in) and 45% with 19 months parallel collection (by the full model). If there is positive correlation between treatment and control groups, the power becomes greater if parallel collection is used but it stays the same without parallel collection[7].

[6] *Note that for phase-in we simulated series with rho=0 and we used κ^{PI} instead of κ^{PR} in initial covariance matrix for SE in the full model*

[7] *Note that correlation does not play any role for phase-in as there are no overlapping treatment and control samples*

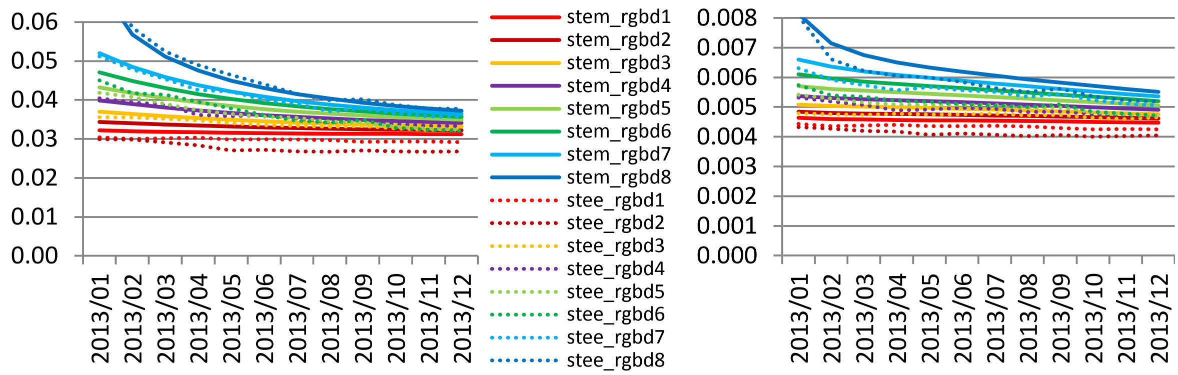
6.3 Simulated (dotted line) and estimated (solid lines) impact after phase-in[8], wave level

a) unemployment b) employment



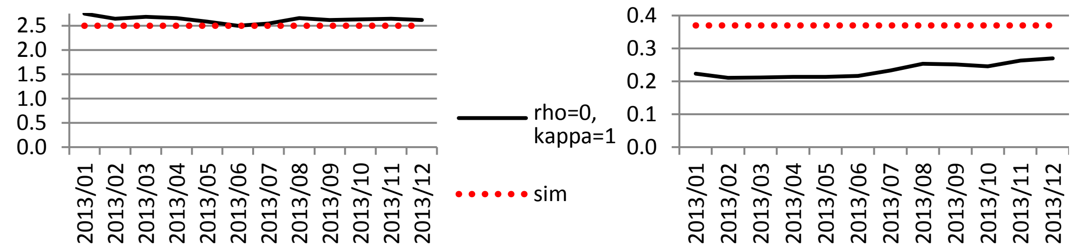
6.4 Empirical (dotted line) and modelled (solid line) SE of impact after phase-in, wave level

a) unemployment b) employment



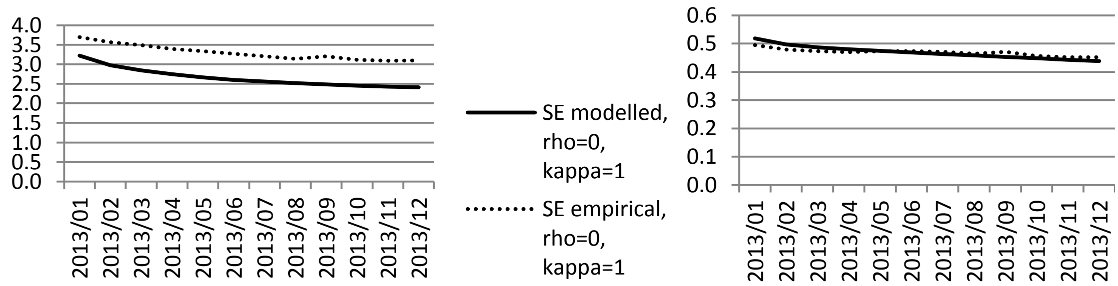
6.5 Simulated (dotted line) and estimated (solid lines) overall impact after phase-in

a) unemployment, % b) employment, %



6.6 Empirical (dotted line) and modelled (solid line) SEs of overall impact after phase-in

a) unemployment, p.p. b) employment, p.p.



[8] *The beginning month in the plots is the eighth month since the beginning of phase-in (when all eight waves have been phased-in).*

6.3 Power (%) of an overall impact detection (full model) 19 months since phase-in commencement

a) empirical

	unemployment	employment
rho	0	0
kappa	1	1
1 SE	12.4	12.7
1.5 SE	22.6	23.2
2 SE	36.4	37.4
2.5 SE	52.2	51.8
3 SE	67.6	67.5

b) modelled

	unemployment	employment
rho	0	0
kappa	1	1
1 SE	17.8	13.2
1.5 SE	34.3	24.4
2 SE	54.6	39.3
2.5 SE	73.7	54.2
3 SE	87.5	70.1

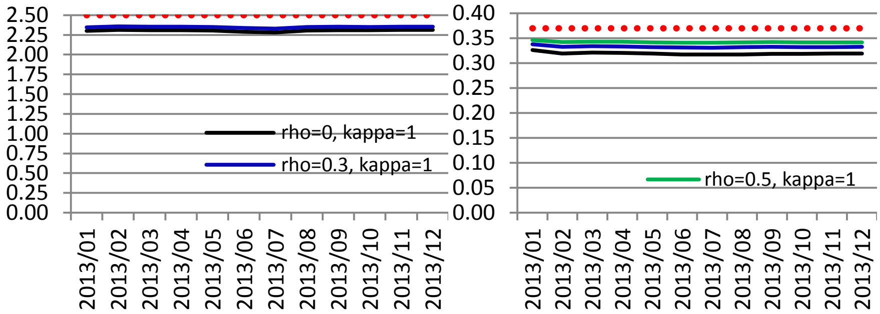
2) Case with parallel collection

If there is parallel collection, there are two models that can be used for impact measurement after parallel collection period – the phase-in model with KF initialisation of impact states estimated from the difference model, and the full model. Because the difference model does not performs as well as the full model during parallel collection period, then the phase-in model with KF initialisation is not a good alternative of the full model for phase-in as well. Further we consider the full model performance only. Also we limit results only to the case with zero correlation and both kappas equal to one. Note that we use parallel collection and phase-in scheme presented on Figure 5.1.

Estimated impacts and their SEs at the wave level results are presented in the Appendix 4 and overall results are presented in Figures 6.5-6.6. The results show that the only improvement in impact SEs is during phase-in period and almost no improvement after phase-in. As standard errors are hardly improving after phase-in the power of impact detection is not increasing as well.

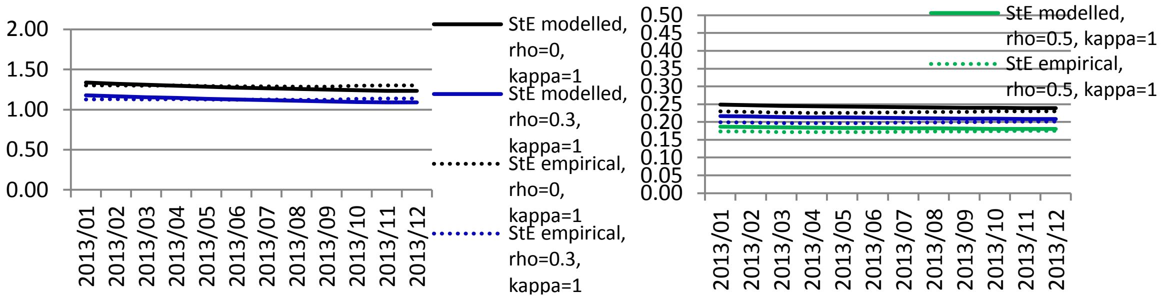
6.7 Simulated (dotted line) and estimated (solid lines) overall impact after phase-in

a) unemployment, % b) employment, %



6.8 Empirical (dotted line) and modelled (solid line) SE of overall impact after phase-in

a) unemployment, p.p. b) employment, p.p.



7. DISCUSSION

In this paper we have presented an integrated approach using multivariate time series models in state-space form that can incorporate a period of parallel collection with a phase-in of the new sample design. We have fully specified the initial conditions for the sampling error process within the KF and demonstrated that achieving high power to detect the desired size of impact will be difficult without a considerable period of parallel collection.

However, while overall power to detect an impact of 1 SE does not reach 80% for practical options, the integrated approach based on the full model delivers important gains in power relative to the difference model during parallel collection and consequently the phase-in period with impact state initialisation. The gain derives from the model using the historical data to remove the common time series effects, rather than just differencing during parallel collection. It is likely that further small gains would be possible using a longer series. This integrated approach also allows for further gains to be explored when additional auxiliary series are available that share correlated components with the series estimated from the survey.

A second advantage derives from the fully integrated nature of the full model. Parallel collection and phase-in are not separate models, and 'initialisation' follows naturally from the end of the parallel collection into the phase-in period. This flexibility allows there to be either a gap between parallel collection and phase-in or a seamless transfer. With the difference model, parallel collection must finish to allow the calculation of all eight effects for initialisation of the phase-in period.

The work presented here has focused on estimating impact at the national level. While achieving a robust measure of an impact at this level is crucial, should an impact be detected it will need to be removed from the published estimates at both national and lower levels. Therefore, the next phase of development will focus on implementing the approach on time series at a sub-group level, with a benchmarking constraint to the national series. Such an approach is implemented by Bureau of Labour Statistics (USA) to produce their sub-national labour market estimates. Further work, not presented here, is developing an approach utilising constrained optimisation techniques to ensure adjusted series at the lowest levels remain consistent with the levels at which we can estimate impact directly.

ACKNOWLEDGEMENTS

The authors are particularly thankful to Dr. Siu-Ming Tam, Sybille McKeown, Prof. Rob Hyndman, Prof. Peter Green, Prof. Ray Chambers, Prof. James Brown, Bjorn Jarvis, Ruel Abello, Steve Xu and MAC reviewer Sc. Prof. Robert Kohn for their constructive and valuable comments and suggestions. The views expressed in the paper do not necessarily represent those of the ABS. Remaining errors are the authors'.

REFERENCES

Australian Bureau of Statistics (ABS) (2018) *Labour Force, Australia, April 2018*, cat. no. 6202.0, ABS, Canberra. <http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/6202.0Explanatory%20Notes1Apr%202018?OpenDocument>

– (2014) *Labour Force, Australia, October 2014*, cat. no. 6202.0, ABS, Canberra. <http://www.abs.gov.au/AUSSTATS/abs@.nsf/Previousproducts/6202.0Main%20Features4Oct%202014?opendocument&tabname=Summary&prodno=6202.0&issue=Oct%202014&num=&view=>

Bailar, B.A. (1975) The effects of rotation group bias on estimates from panel surveys. *Journal of the American Statistical Association*, Vol. 6, (Eds, P.R. Krishnaiah and C.R. Rao), Amsterdam: Elsevier Science, 187-211.

Binder, D.A. and Hidiroglou, M.A. (1988) Sampling in time. In *Handbook of Statistics*, 70, 23-30.

Durbin, J. and Koopman, S.J. (2012) *Time series analysis by state space methods*. Oxford: Oxford University Press.

Kumar, S. and Lee, H. (1983) Evaluation of composite estimation for the Canadian labour force survey. *Survey Methodology*, 9, 178-201.

Pfeffermann, D. (1991) Estimation and seasonal adjustment of population means using data from repeated surveys. *Journal of Business & Economic Statistics*, 9, pp. 163-175.

Pfeffermann, D., Feder, M. and Singnorelli, D. (1998) Estimation of autocorrecations of survey errors with application to trend estimation in small areas. *Journal of Business and Economic Statistics*, Vol. 16, No. 3 (Jul., 1998), pp. 339-348.

Van den Brakel, J. and Krieg, S. (2009) Estimation of the monthly unemployment rate through structural time series modelling in a rotating panel design. *Survey Methodology*, 35, pp. 177-190.

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

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

Van den Brakel, J., Smith, P.A. and Compton, S. (2008) Quality procedures for survey transitions, experiments, time series and discontinuities. *Journal for Survey Research Methods*, 2(3), pp. 123-141.

Zhang, M. and Honchar, O. (2016) Predicting survey estimates by state space models using multiple data sources. *Methodology Advisory Committee Meeting, Australian Bureau of Statistics. June 2016*

Zhang, M., Van den Brakel, J., Honchar, O., Wong, C. and Griffiths, G. (2018) Using state space models for measuring statistical impacts of survey redesigns – A case study of the ABS labour force survey. *Research paper, Australian Bureau of Statistics.*

APPENDIXES

Appendix 1. Sampling error variance and autocorrelation calculation

There were two methods used for SE autocorrelation calculation in this study: pseudo-errors method developed by Pfeffermann *et.al.* (1998) and GREG weighted residuals method. SE variance was estimated using weighted residuals method. Below we describe these two methods in more detail.

Pseudo-errors method

1. Calculate GREG estimates for number of employed people and number of unemployed people by rotation groups monthly for period from June 2006 till May 2016. GREG estimates have to be obtained by calibrating independently each rotation group to overall population benchmarks.
2. To each rotation group assign respective time-in-survey (wave, panel) j.
3. For each month within the time period calculate average estimate y_t (for employment and unemployment) across rotation groups by formula $y_t = \sum_{j=1}^{8} y_t^{(j)} \big/ 8$, where 8 is the number of rotation groups in LFS.
4. For each panel calculate pseudo panel-survey errors $e_{tp}^{(j)} = y_t^{(j)} - y_t$, where $y_t^{(j)}$ is a GREG estimate (of employment or unemployment) for panel j at month t.
5. For each panel calculate average: $e_p^{(j)} = \frac{1}{N}\sum_{t=1}^{N} e_{tp}^{(j)}$, where N is the number of months in the time period.
6. For each panel j calculate averages at different lags k ($k=1,...10$) : $e_p^{j,k} = \frac{1}{N}\sum_{t=k+1}^{N+k} e_{t-k,p}^{j,t}$, where $e_{t-k,p}^{j,t}$ is the pseudo error at month *(t-k)* corresponding to the panel selected from the same rotation group as the panel enumerated for the j th time at month t (in particular, $e_{t,p}^{j,t} = e_{tp}^{(j)}$).
7. For each panel calculate covariances $\hat{C}_k^j$ and variances $\hat{C}_0^j$ at various lags (including $k=0$ for variance), $k=0,1,...10$: $\hat{C}_k^j = \frac{1}{N}\sum_{t=k+1}^{N}\left(e_{tp}^{(j)} - e_p^{(j)}\right)\left(e_{t-k,p}^{j,t} - e_p^{j,k}\right)$.
8. Calculate sampling error autocorrelations at the national level[9] at lags $k=1,...10$: $\hat{\rho}_k = \sum_{j=1}^{8}\hat{C}_k^j \Big/ \sum_{j=1}^{8}\hat{C}_0^j$.
9. Calculate sampling error autocorrelations at the panel level at different lags:

[9] *Note overall (national level) sampling error autocorrelations and variances were used in simulation study to formulate "real world component" in simulated series by waves.*

$$\hat{\rho}_k^j = \frac{\sum_{t=k+1}^{N}\left(e_{tp}^{(j)} - e_p^{(j)}\right)\left(e_{t-k,p}^{j,t} - e_p^{j,k}\right)}{\sqrt{\sum_{t=1}^{N}\left(e_{tp}^{(j)} - e_p^{(j)}\right)^2}\sqrt{\sum_{t=k+1}^{N+k}\left(e_{t-k,p}^{j,t} - e_p^{j,k}\right)^2}} .$$

GREG weighted residuals method

In calculation the following steps were taken:

1. For calculation of SE variance by rotation groups the unit-level LFS data files need to be split by rotation groups and calculation can be then conducted for each rotation group separately following steps described below.

   - Calculate individual weighted residuals for each person on the file

     $w_i(y_i - x_i\beta)$,

     where w_i is a GREG weight, y_i is a labour force status (employed, unemployed) for person i, $x_i\beta$ is a GREG prediction for person i obtained from calibration to overall benchmarks (population totals by age, sex etc.)

   - Calculate sum of weighted residuals for each variance group g in each stratum h by labour force status (employed, unemployed): $e_{hg} = \sum_{i\in hg} w_i\left(y_i - x_i\beta\right)$.

     The variance groups were assigned systematically to the sorted list of Primary Sampling Unit blocks. There were 32 variance groups used for estimation of overall sampling error variance and 20 variance groups for SE variance estimation by rotation groups (i.e. 20 in each rotation group).

   - Calculate sum of weighted residuals for each stratum h divided by number of variance groups in a stratum by labour force status: $e_h = \sum_{i\in h} w_i\left(y_i - x_i\beta\right)/32$.

     Note there will be 20 instead of 32 for variance calculation by rotation group (here and in following steps).

   - Calculate weighted residuals variance for each month in calculation: $Var(\hat{y}) = \sum_h (31/32)\sum_{g\in h}\left(e_{hg} - e_h\right)^2$.

2. For each pair of months at the various lags within the time period calculate:

   - variance of movement between pair of months at various lags (lag k =1,..,10): $Var(\hat{y}^1 - \hat{y}^2)$,

   - covariance between pair of months at various lags: $Cov(\hat{y}^1, \hat{y}^2) = 1/2\left(Var\left(\hat{y}^1\right) + Var\left(\hat{y}^2\right) - Var\left(\hat{y}^1 - \hat{y}^2\right)\right)$,

   - correlation between pair of months at lag k: $Corr(\hat{y}^1, \hat{y}^2) = \dfrac{Cov\left(\hat{y}^1, \hat{y}^2\right)}{\sqrt{Var\left(\hat{y}^1\right)\cdot Var\left(\hat{y}^2\right)}}$.

Appendix 2. Sampling error simulation

The structure of the errors in the wave level time series is complex because the AR(2) structure does not operate within a wave but as a lagged dependency across waves. For the first rotation group, we then create the following sequence of sampling errors:

$$e_{11} \sim N(0, \sigma^2)$$
$$e_{22} = \delta_{11} e_{11} + u_{22}, \qquad u_{22} \sim N(0, \sigma_2^2)$$
$$e_{33} = \delta_{21} e_{22} + \delta_{22} e_{11} + u_{33}, \qquad u_{33} \sim N(0, \sigma_3^2)$$
$$e_{44} = \delta_{21} e_{33} + \delta_{22} e_{22} + u_{44}, \qquad u_{44} \sim N(0, \sigma_4^2)$$
$$\cdots$$
$$e_{88} = \delta_{21} e_{77} + \delta_{22} e_{66} + u_{88}, \qquad u_{88} \sim N(0, \sigma_8^2)$$

$$e_{91} \sim N(0, \sigma^2)$$
$$e_{10,2} = \delta_{11} e_{91} + u_{10,2}, \qquad u_{10,2} \sim N(0, \sigma_2^2)$$
$$e_{11,3} = \delta_{21} e_{10,2} + \delta_{22} e_{9,1} + u_{11,3}, \qquad u_{11,3} \sim N(0, \sigma_3^2)$$
$$\cdots$$

where σ^2 is the wave level sampling variance for the LFS, the σ_i's are the parameters that define the AR(2) process within a rotation group, and the σ_i^2's are the appropriate white noise disturbance variances to ensure the overall variance of e_{ti} is always the estimated sampling variance σ^2. Note that these disturbance variances need to be derived as we do not have a standard stable AR(2) process, it re-starts with a new uncorrelated error after every eight time-points. However, σ_8^2 is very close to the long-term disturbance error variance in a long-sequence AR(2) process.

For the second rotation group, we then create the following sequence of errors independent of the first rotation group:

$$e_{21} \sim N(0, \sigma^2)$$
$$e_{32} = \delta_{11} e_{21} + u_{32}, \qquad u_{32} \sim N(0, \sigma_2^2)$$
$$e_{43} = \delta_{21} e_{32} + \delta_{22} e_{21} + u_{43}, \qquad u_{43} \sim N(0, \sigma_3^2)$$
$$e_{54} = \delta_{21} e_{43} + \delta_{22} e_{32} + u_{54}, \qquad u_{54} \sim N(0, \sigma_4^2)$$
$$\cdots$$
$$e_{98} = \delta_{21} e_{87} + \delta_{22} e_{76} + u_{98}, \qquad u_{98} \sim N(0, \sigma_8^2)$$
$$e_{10,1} \sim N(0, \sigma^2)$$
$$e_{11,2} = \delta_{11} e_{10,1} + u_{11,2}, \qquad u_{11,2} \sim N(0, \sigma_2^2)$$
$$e_{12,3} = \delta_{21} e_{11,2} + \delta_{22} e_{10,1} + u_{12,3}, \qquad u_{12,3} \sim N(0, \sigma_3^2)$$
$$\cdots$$

utilising the same parameters as for the first rotation group, but starting with first time in survey at time 2. We repeat this independent generation up to the eighth rotation group, with the following sequence of errors:

$$e_{81} \sim N(0, \sigma^2)$$

$$e_{92} = \delta_{11} e_{81} + u_{92}, \qquad u_{92} \sim N(0, \sigma_2^2)$$
$$e_{10,3} = \delta_{21} e_{92} + \delta_{22} e_{81} + u_{10,3}, \qquad u_{10,3} \sim N(0, \sigma_3^2)$$
$$e_{11,4} = \delta_{21} e_{10,3} + \delta_{22} e_{92} + u_{11,4}, \qquad u_{11,4} \sim N(0, \sigma_4^2)$$
$$\ldots$$
$$e_{15,8} = \delta_{21} e_{14,7} + \delta_{22} e_{13,6} + u_{15,8}, \qquad u_{15,8} \sim N(0, \sigma_8^2)$$
$$e_{16,1} \sim N(0, \sigma^2)$$
$$e_{17,2} = \delta_{11} e_{16,1} + u_{17,2}, \qquad u_{17,2} \sim N(0, \sigma_2^2)$$
$$e_{18,3} = \delta_{21} e_{17,2} + \delta_{22} e_{16,1} + u_{18,3}, \qquad u_{18,3} \sim N(0, \sigma_3^2)$$
$$\ldots$$

Once we have the errors as sequences in rotation group structure, we can re-structure the errors into the wave form

$$\begin{array}{llllllll}
e_{11} \\
e_{21} & e_{22} \\
e_{31} & e_{32} & e_{33} \\
e_{41} & e_{42} & e_{43} & e_{44} \\
e_{51} & e_{52} & e_{53} & e_{54} & e_{55} \\
e_{61} & e_{62} & e_{63} & e_{64} & e_{65} & e_{66} \\
e_{71} & e_{72} & e_{73} & e_{74} & e_{75} & e_{76} & e_{77} \\
e_{81} & e_{82} & e_{83} & e_{84} & e_{85} & e_{86} & e_{87} & e_{88} \\
e_{91} & e_{92} & e_{93} & e_{94} & e_{95} & e_{96} & e_{97} & e_{98}
\end{array}$$
$$\ldots$$

so that the first wave is first-time-in-survey for each rotation group, second wave is second-time-in-survey, and so on.

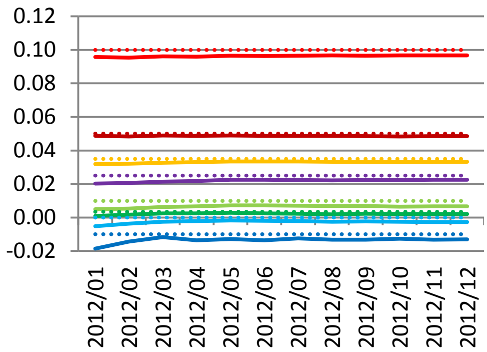
We repeat the error generation process for $\tilde{e}_{ti}$ using the same AR parameters but using the sampling variance $\tilde{\sigma}^2$. The $\tilde{e}_{ti}$ are generated such that their correlation with e_{ti} is φ.

Appendix 3. Impact estimates and their SEs during parallel collection period, wave level[10]

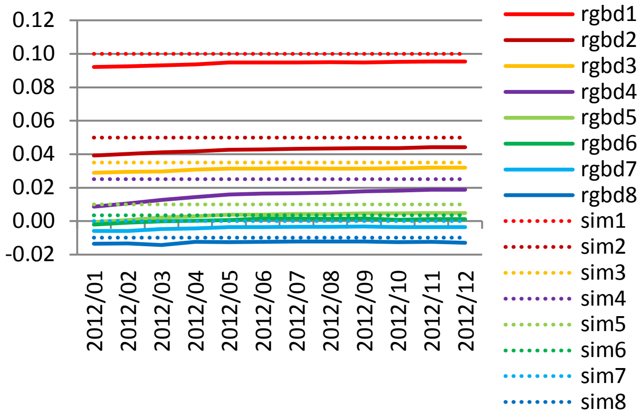
1) Unemployment

A3.1 Simulated (dotted line) and estimated (solid line) impact

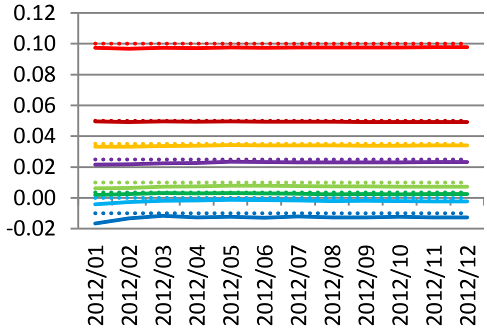
a) full model, rho=0, kappa=1



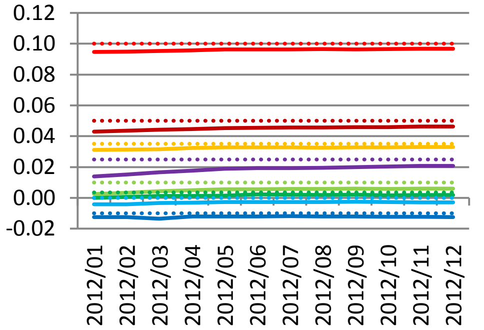
b) difference model, rho=0, kappa=1



c) full model, rho=0.3, kappa=1

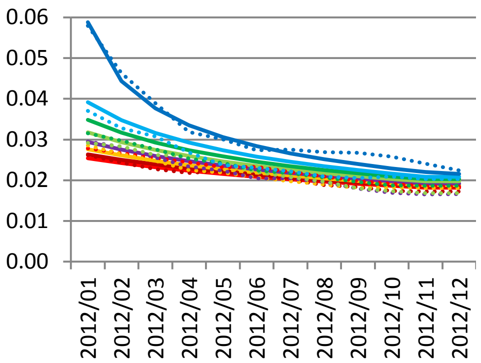


d) difference model, rho=0.3, kappa=1

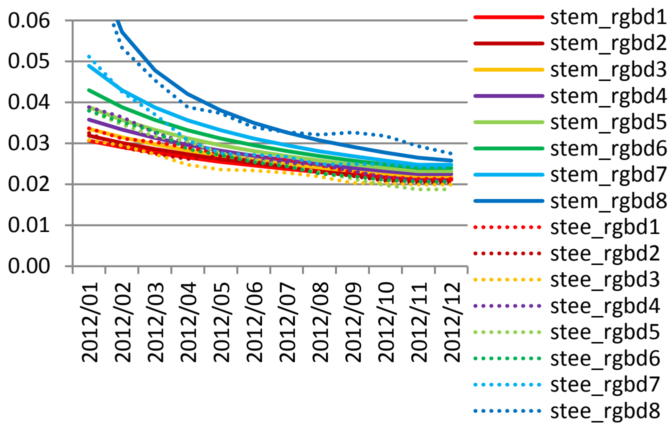


A3.2 Empirical (dotted line) and modelled (solid line) impact SE

a) full model, rho=0, kappa=1



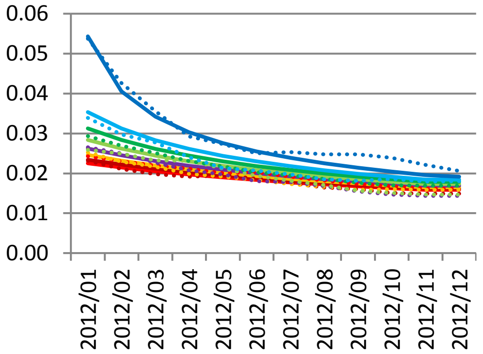
b) difference model, rho=0, kappa=1



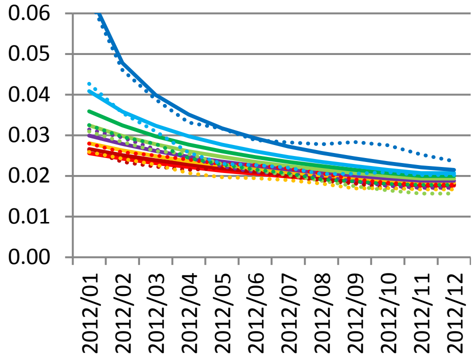
---

[10] *The beginning month at the plots is the eighth month since beginning of parallel collection period (when all eight waves have been phased-in).*

c) full model, rho=0.3, kappa=1



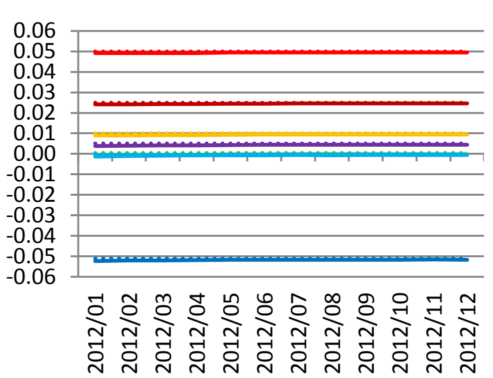
d) difference model, rho=0.3, kappa=1



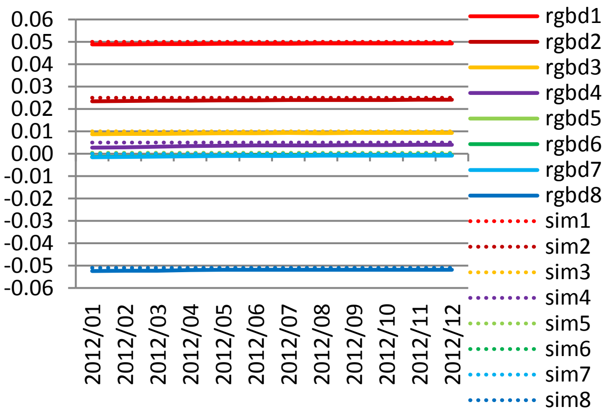
2) Employment

A3.3 Simulated (dotted line) and estimated (solid line) impact

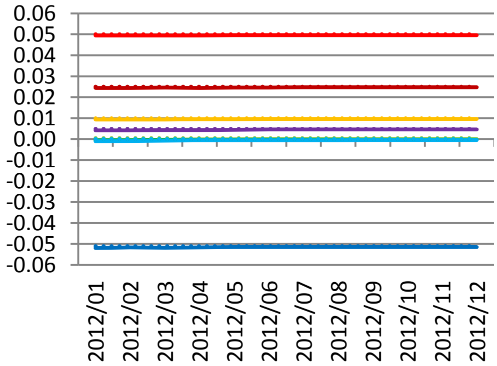
a) full model, rho=0, kappa=1



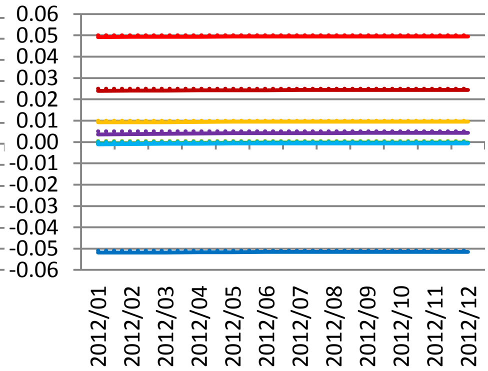
b) difference model, rho=0, kappa=1



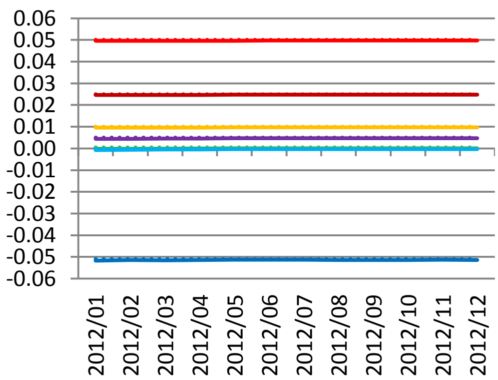
c) full model, rho=0.3, kappa=1



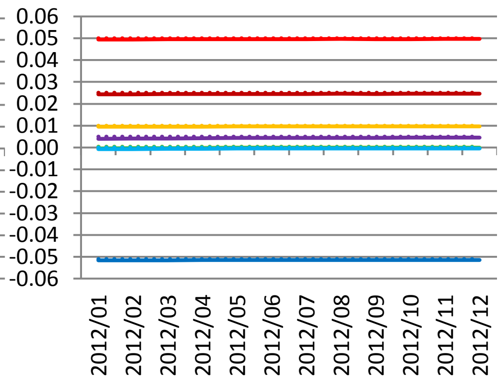
d) difference model, rho=0.3, kappa=1



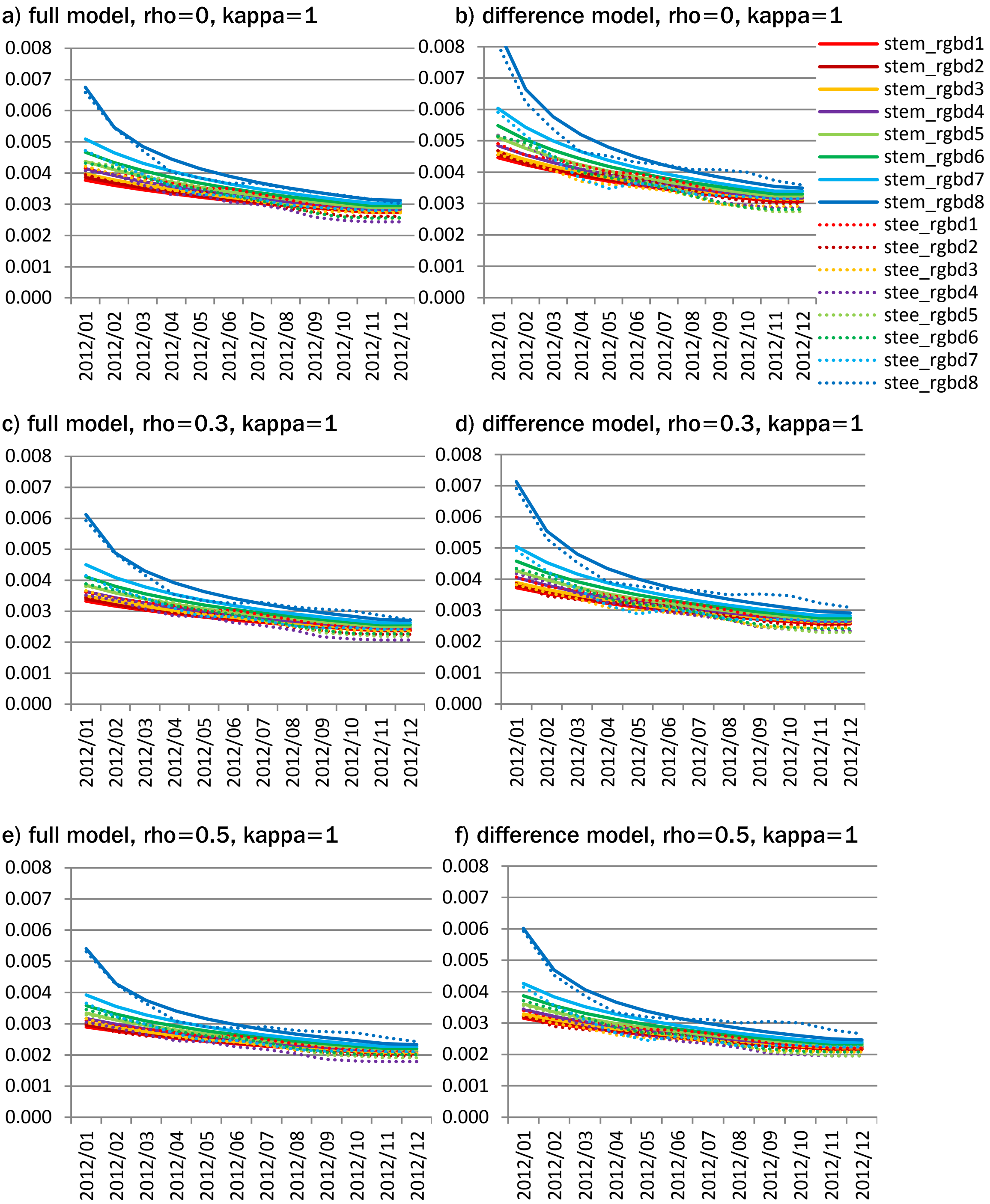
e) full model, rho=0.5, kappa=1



f) difference model, rho=0.5, kappa=1



A3.4 Empirical (dotted line) and modelled (solid line) impact SE

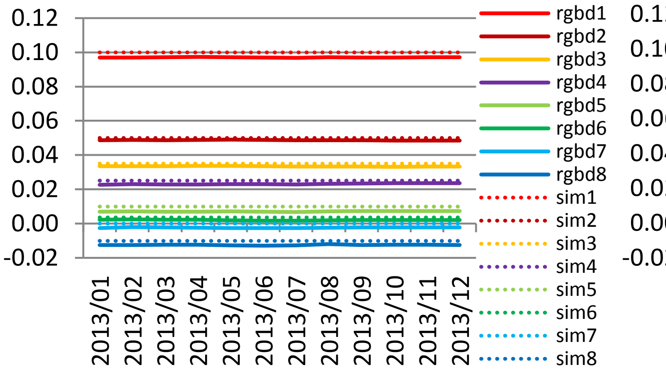


Appendix 4. Impact estimates and their SEs after phase-in, wave level[11]

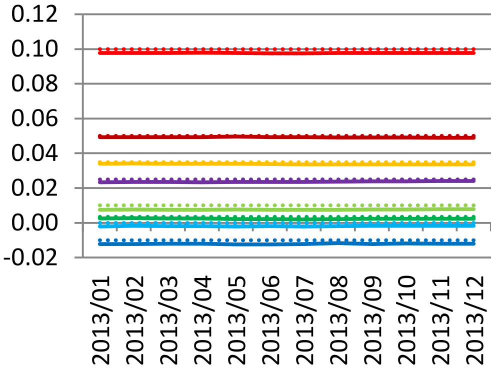
1) Unemployment

A4.1 Simulated (dotted line) and estimated (solid line) impact

a) full model, rho=0, kappa=1

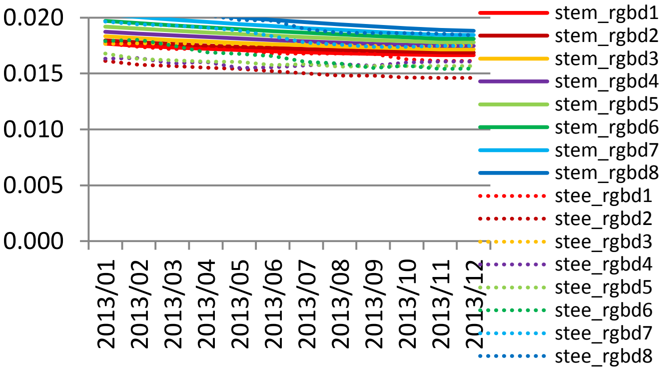


b) full model, rho=0.3, kappa=1

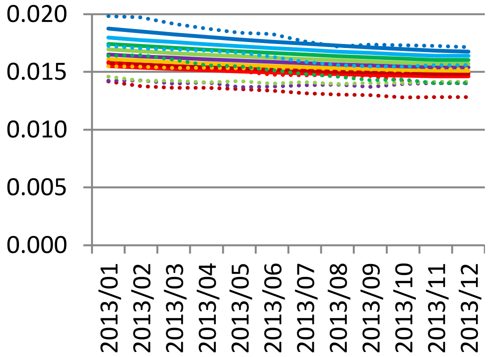


A4.2 Empirical (dotted line) and modelled (solid line) impact SE

a) full model, rho=0, kappa=1



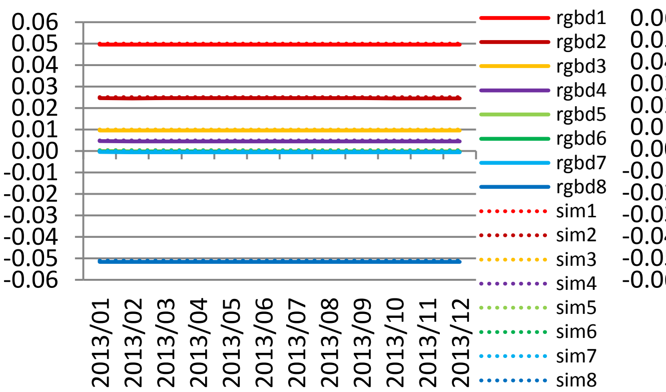
b) full model, rho=0.3, kappa=1



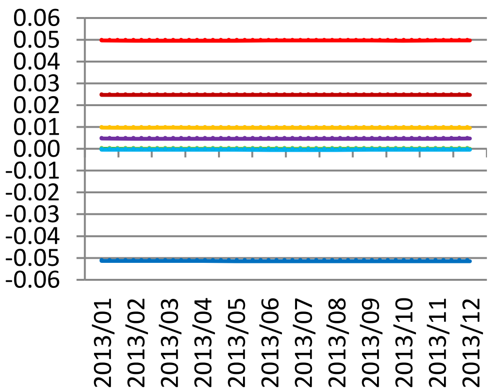
2) Employment

A4.3 Simulated (dotted line) and estimated (solid line) impact

a) full model, rho=0, kappa=1

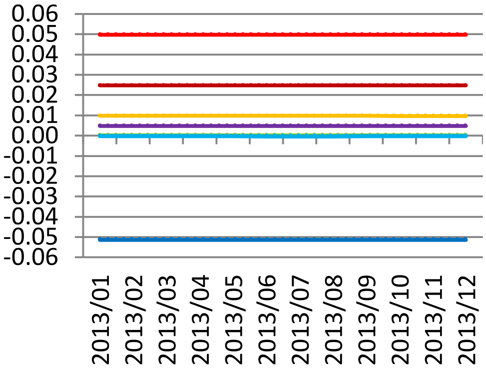


b) full model, rho=0.3, kappa=1



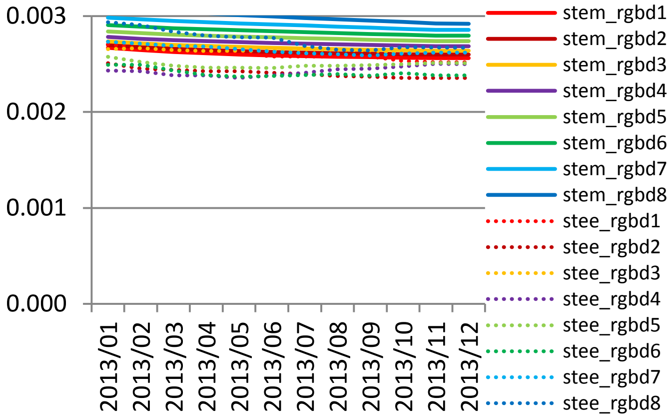
[11] *The beginning month at the plots is the eighth month since beginning of phase-in period (when all eight waves have been phased-in).*

c) full model, rho=0.5, kappa=1

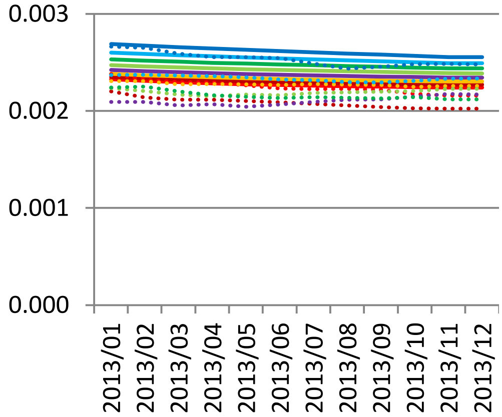


A4.4 Empirical (dotted line) and modelled (solid line) impact SE

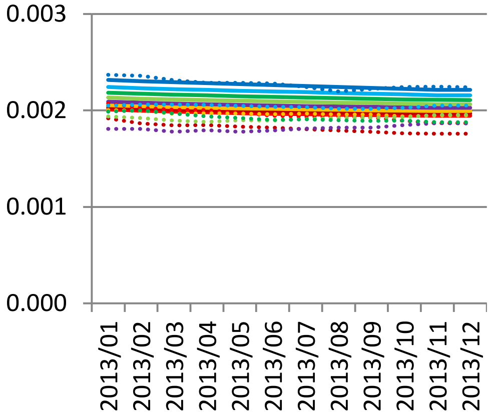
a) full model, rho=0, kappa=1



b) full model, rho=0.3, kappa=1



c) full model, rho=0.5, kappa=1



FOR MORE INFORMATION . . .

www.abs.gov.au the ABS website is the best place for data from our publications and information about the ABS.

INFORMATION AND REFERRAL SERVICE

Our consultants can help you access the full range of information published by the ABS that is available free of charge from our website. Information tailored to your needs can also be requested as a 'user pays' service. Specialists are on hand to help you with analytical or methodological advice.

POST Client Services, ABS, GPO Box 796, Sydney NSW 2001

FAX 1300 135 211

EMAIL client.services@abs.gov.au

PHONE 1300 135 070

FREE ACCESS TO STATISTICS

All ABS statistics can be downloaded free of charge from the ABS web site.

WEB ADDRESS www.abs.gov.au

Produced by the Australian Bureau of Statistics