Australian Bureau of Statistics

1351.0.55.160

Research Paper

Using State Space Models for Measuring Statistical Impacts of Survey Redesigns - A case study of the ABS Labour Force Survey

Australia

2018

AUSTRALIAN BUREAU OF STATISTICS

Research Paper

Using State Space Models for Measuring Statistical Impacts of Survey Redesigns

- A case study of the ABS labour force survey

Xichuan (Mark) Zhang[1], Jan Van Den Brakel[2],
Oksana Honchar[1], Cedric Wong[1] and Greg Griffiths[1]

1. Methodology Division, Australian Bureau of Statistics. 2. Department of Statistical Methods, Statistics Netherlands and Department of Quantitative Economics, Maastricht University School of Business and Economics

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 11.30AM (CANBERRA TIME) WED 21 MARCH 2018

ABS Catalogue No. 1351.0.55.160

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.

CONTENTS

ABBREVIATIONS

ABS	Australian Bureau of Statistics
AR	Auto Regressive
GREG	Generalised Regression
LS	Level Shift
LFS	Labour Force Survey
MDI	Minimum Detectable Impact
PSU	Primary Sample Unit
RBD	Randomized Block Design
RGB	Rotation Group Bias
SIM	Statistical Impact Measurement
SSM	State Space Model
STM	Structural Time Series Model

USING STATE SPACE MODELS FOR STATISTICAL IMPACT MEASUREMENT OF SURVEY REDESIGNS

- A CASE STUDY OF ABS LABOUR FORCE SURVEY

Xichuan (Mark) Zhang[1], Jan Van Den Brakel[2],
Oksana Honchar[1], Cedric Wong[1] and Greg Griffiths[1]

ABSTRACT

The implementation of a major business transformation program in an official statistical agency is often designed to achieve, among other things, improvements in data collection efficiency, data processing methodology and data quality. However, achieving such improvements can, in itself, result in transitional statistical impacts which could be misinterpreted as real world change if they are not measured and handled appropriately.

This paper describes early work to explore a range of statistical methods for measuring the statistical impacts which could be encountered in a survey redesign, using the ABS Labour Force Survey (LFS) as a case study, including:

1. designing experiments for field trials of different questionnaires and data collection strategies;
2. designing and conducting parallel collection activities such that the outgoing and the incoming surveys are run in parallel for a period of time to measure the impact of any collection changes; and
3. refining the precision of impact measurement while implementing a new survey design.

The results presented are for illustrative purposes for further development rather than for an actual implementation to the Australian LFS.

State space modelling techniques have been utilised as the main approach for efficient impact measurement. This approach enables us to incorporate sampling error structure and time series intervention. The approach can also be extended to take advantage of other related data sources to improve impact measurement efficiency and accuracy. While the LFS is used as a case study, the models and methods developed can be extended to other surveys.

[1] Methodology Division, Australian Bureau of Statistics.
[2] Department of Statistical Methods, Statistics Netherlands and Department of Quantitative Economics, Maastricht University School of Business and Economics

1. INTRODUCTION

It is common practice for national statistical offices to employ a repeated sampling scheme. This enables estimation of changes for the total aggregate (or population) as well as cross-sectional estimates. The time series produced under the repeated survey scheme over time create a basis for social, economic, environmental analysis and policy making.

Any changes in survey methodology would potentially affect the continuity of the estimated time series. This creates difficulties for users in interpreting movements in data and making policy decisions, because it may not be clear if the unusual movements in the estimates represent real world changes or if they are measurement changes introduced by new or alternative methodological approaches. Therefore any changes in survey methodology have to be well managed, the impact of methodological change need to be identified, measured and adjusted, if necessary, to provide a coherent picture before and after the change and to mitigate the risk of misinterpretation of the changes.

The Australian Bureau of Statistics (ABS) is embarking on a transformation program, which includes, amongst other changes, applying different collection modes for survey data and using different, but more efficient, sampling frames and estimation methods for official statistics. Whilst this transformation is expected to deliver positive changes to official statistics, there is a risk that such changes could have a statistical impact on some ABS time series. The challenge is to develop methodologies to measure, and where needed adjust for such statistical impacts. A general frame work for statistical impact measurement is described by Van den Brakel et al. (2017)

The first and the most straightforward approach to assess impact of survey changes is to conduct a parallel run, i.e. to conduct the survey under the old and new approach simultaneously (see, for example, Van den Brakel, 2008). The outgoing and the incoming surveys are run in parallel for a period of time in order to collect information about any impact of the change.

Various intervention analyses of time series models are also widely utilised to measure the possible time series discontinuities with and without utilising the prior information from a parallel run. For example, Van den Brakel and Krieg (2015) describe how a multivariate structural time series model was used to measure statistical impact induced by the Dutch Labour Force survey redesign.

The challenge for measuring statistical impacts is managing the trade-off among different priorities including:

- statistical accuracy required (minimum detectable impact and both type I and II errors)
- operational feasibility and impacts, and

- Cost

with consideration to:

- target statistics
- the assumption of the intra cluster correlation between the control[3] and treatment samples
- desired dissection of changes to be measured – net or component impacts
- appetite for accepting more volatile published estimates during the parallel period if sample size is reduced
- appetite for accepting revisions after the new survey implementation due to the uncertainty of the measured impact derived from small treatment sample.

The ABS is establishing a three-phase statistical impact measurement (SIM) strategy:

Phase 1: Experimental design and field tests for measuring the effectiveness and broad statistical impact of a change and for making decisions about the final design of the new survey process.

Phase 2: A parallel run approach to collect data and measure statistical impacts with required accuracy.

Phase 3: Implementation of changes, adjustment for statistical impacts, monitoring of the change process, and revision if necessary.

This paper primarily presents initial development work for Phases 2 and 3 using state space modelling techniques to handle some special characteristics of the ABS Labour Force Survey (LFS), such as:

- changes that are rotation group wave sensitive, and
- the smoothing effect of composite estimates,

while also considering:

- the need to balance different priorities, and
- the utilisation of statistical impact information from previous phases for further improvement.

The results presented are for illustrative purposes for further development rather than for an actual implementation to the Australian LFS.

Section 2 provides a brief introduction of the characteristics of the current ABS LFS survey and possible future changes. It also includes general structural time series models and their state space presentation for measuring statistical impact. Section 3 describes the methods and models that could be used for LFS parallel run design and discusses simulated results. Section 4 presents a method to improve State Space Model (SSM) performance for short time series, and provides an alternative

[3] Control samples are also referred as regular samples to produce the regular estimates for publication.

embedded experimental design approach for further study. Section 5 evaluates some options under consideration by the ABS, and suggests a hybrid option to balance different priorities in terms of cost, accuracy and revisions. Section 6 discusses the implications of different options and future work to support the three phase SIM strategy.

All the calculations reported in this paper were carried out with programs written in the SSM procedure in SAS, SsfPack (see Koopman *et al*, 2008) and R.

2. ABS LABOUR FORCE SURVEY

2.1 ABS Labour Force Survey

The Labour Force Survey (LFS) is based on a multi-stage area sample of dwellings and covers approximately 0.32% of the civilian population of Australia aged 15 years and over (ABS, 2016). Households selected for the LFS are interviewed using face-to-face, phone or web form each month for eight consecutive months, with one-eighth of the sample being replaced each month. The LFS sample can be thought of as comprising eight sub-samples (or rotation groups, RG hereafter), with each sub-sample remaining in the survey for eight months, and one rotation group "rotating out" each month and being replaced by a new group "rotating in". This high overlap of respondents from month-to-month induces a strong serial correlation into the sampling errors. In addition, the replacement sample is generally selected from the same geographic areas as the outgoing one, as part of a representative sampling approach. This induces additional serial correlation in the sampling errors even for non-overlapping outgoing and incoming sub-samples.

The estimation method used in the LFS is composite estimation. By exploiting the high correlation between overlapping samples across months, the ABS LFS composite estimator combines the previous six and current months' data by applying different factors (also called BLUE multipliers) according to length of time in the survey (or waves. e.g. wave 1 is the first time in survey etc.). After these factors are applied seven months of data are weighted to align with current month population benchmarks.

2.2 Reasons for measuring statistical impact using GREG estimates at rotation group level

Although our interest is to measure a statistical impact at the level of composite estimates, there are different ways to make measurement at different levels. In order to achieve accurate and timeliness measurement, we propose to measure the statistical impact at rotation group level because of the property of the composite estimator.

The multipliers of the ABS LFS composite estimator are applied to past observations of GREG estimates at rotation group level to produce current LFS estimates. As a result, a smoothing effect will apply to any abrupt statistical impact at the current end of the series. To avoid such an effect, SIM modelling work needs to be conducted on the GREG estimates, i.e. prior to applying the composite multipliers, so that detection and measurement of abrupt impacts is timely and accurate. The corresponding impacts to the composite LFS estimates can then be derived accordingly.

There are a number of potential changes to the LFS which must be considered and their statistical impacts assessed. It is not realistic to assume a statistical impact is uniformly equal to all waves because the proposed changes may have different impacts on different waves. We refer to this as "wave sensitive" in this paper.

The following are examples of possible future changes which could be wave sensitive.

- Use of e-collection as the primary collection mode. Changes to respondent induction and the strategy for promoting web-form take-up potentially lead to a wave sensitive effect.
- Changing the placement and timing of supplementary surveys in the monthly population survey of which the LFS is a part. This may be wave sensitive because supplementary survey placement is currently wave dependent.
- Implementing responsive design follow-up, to ensure data acquisition effort is focused on quality samples and is cost-effective. This can potentially have more impact on wave one versus other waves.
- Further wave dependent effects are possibly driven by any wave dependence in response profiles.

2.3 Structural time series model at the rotation group level for measuring statistical impact

Assume $\hat{y}_{i,t}$ is a GREG estimate of the main LFS variables such as number of employed and number of unemployed persons from the rotation group that in the current month t has been observed i times ($i=1, \ldots, 8$) (referred to as wave i hereafter). Without losing generality, the structural measurement errors for the wave i at time t are

1. rotation group bias (RGB) b_i , and
2. sampling error $e_{i,t}$.

The rotation group bias $b_1,...,b_8$ reflects a permanent wave sensitive level shift compared to one reference wave (in this study reference wave was wave 7, therefore $b_7 = 0$).

The following equation[4] describes the relationship between an observed estimate $\hat{y}_{i,t}$ and unobserved components y_t, b_i , $e_{i,t}$ and the intervention effect (permanent level shift) α_i with a time invariant assumption[5] :

$$\begin{pmatrix} \hat{y}_{1,t} \\ \vdots \\ \hat{y}_{8,t} \end{pmatrix} = 1_{[8]} y_t + \begin{pmatrix} b_1 \\ \vdots \\ b_8 \end{pmatrix} + \begin{pmatrix} \alpha_1 x_{1,t} \\ \vdots \\ \alpha_8 x_{8,t} \end{pmatrix} + \begin{pmatrix} e_{1,t} \\ \vdots \\ e_{8,t} \end{pmatrix} \tag{2.1}$$

where y_t is a true population value, $1_{[8]}$ is the 8 dimensional identity matrix and x_i is an intervention dummy variable denoted as

$$x_{i,t} = \begin{cases} 1, & \text{if observations are obtained under the new design of wave } i \text{ at time } t, \\ 0, & \text{otherwise.} \end{cases}$$

y_t can be expressed by a structural time series model (STM)

$$y_t = T_t + S_t + I_t, \tag{2.2}$$

where T_t, S_t, and I_t denote the smooth trend model, seasonal model, and irregular which often is assumed as white noise for unexplained variation of the population parameter, respectively, see Durbin and Koopman (2012) for details. The sampling error stochastic process e_t can be modelled as white noise for wave 1 (assuming no correlation with estimates from previous panel)

$$e_{1,t} = u_{1,t}, \quad u_{1,t} \cong NID\left(0, \sigma^2_{1,u}\right) \tag{2.3}$$

as an AR(1) process for wave two

$$e_{2,t} = \phi_{2,1} e_{1,t-1} + u_{2,t}, \quad u_{2,t} \cong NID\left(0, \sigma^2_{2,u}\right) \tag{2.4}$$

and as an AR(2) process for other waves (i=3,4,…8)

$$e_{i,t} = \phi_1 e_{i-1,t-1} + \phi_2 e_{i-2,t-2} + u_{i,t}, \quad u_{i,t} \cong NID\left(0, \sigma^2_{i,u}\right) \tag{2.5}$$

where coefficients ϕ_1 and ϕ_2 and the sampling error disturbance variance σ^2_u can be predefined from the LFS data (see Pfeffermann et al., 1998).

---

[4] All the modelling work is on logarithmic scale in this paper. Standard Error (SE) is equivalent to Relative Standard Error (RSE) in the original scale. They may be used interchangeably in this paper.

[5] Further elaboration of this simple model may be needed if evidence emerges that this assumption needs revision. See some detailed discussion for future in Section 6

3. PARALLEL RUN DESIGN

Design Considerations in the LFS Context

The objectives of any LFS parallel run design would be to:

- measure the direct statistical impacts induced by the ABS process change to the published ABS LFS outputs rather than unit level impacts,
- identify statistical impacts in a timely manner to support statistical risk management, and
- obtain accurate statistical impact measurement with a minimum treatment sample for the agreed accuracy level, and a feasible parallel run design.

Working assumptions

For this study, the hypothetical accuracy criterion[6] is set to detect a significant statistical impact as follows:

One standard error of population [7] estimates (43750 and 19500 for employed and unemployed respectively) with the conventional Type I and II errors less than 5% and 50% respectively.

The minimum detectable impact (MDI) is defined as the size of the impact that can be detected based on the above stated accuracy criterion. Its value is calculated as the standard error of the estimated statistical impact times a multiplier which is derived from a set of predefined Type I and II errors. The multiplier is 1.96 for Type I and II errors of 5% and 50% respectively (see Section 2 of Appendix 1). The ratio of MDI to one standard error of the population estimate (referred as MDI ratio hereafter) indicates that a statistical impact measurement method is successful when its value is less than or equal to one. MDI ratio provides a uniform measure and makes the comparisons of SIM for different variables easier.

Scope of the impacts to be measured

The statistical impact measurement described here is primarily designed for identifying a permanent level shift induced by a new LFS design with additional consideration of sampling error properties

The following issues were broadly considered, but put out of scope of this current work:

1. Some aspects of a new LFS design may have short term transitory impacts, e.g. due to unfamiliarity with new operational processes etc. These types of

[6] There is no official accuracy as this paper is written. The hypothetical accuracy is purely for assisting discussion for this paper.

[7] Population is referred as employed or unemployed persons in the context of this paper

transitory impact should be mitigated by early training and better preparation, rather than measured as a part of a parallel run.

2. Some changes may have permanent impacts on seasonal patterns. The affordable size and duration of any parallel run is unlikely to be enough to precisely estimate any such change. The nature of such changes would need to be assessed broadly and qualitatively either prior to parallel run via experiments and field testing or through ongoing monitoring over a longer period. An STM, for example, that allows for a break in the seasonal pattern after the changeover can be used to assess impact on seasonal patterns several years after the changeover.

Parameters in parallel run design

The following two parameters for a parallel run design are required to meet the accuracy criterion and operational feasibility:

(1) the size of the treatment sample; and

(2) the duration of the parallel run.

From an operational feasibility perspective, the duration of any parallel run in this study is limited to less than two years.

3.1 State Space Model formulation

The set of equations (2.1) to (2.5) have described a general SSM framework for GREG estimates of LFS at the rotation group level with interventions. This model is a common approach in the literature to measure a statistical impact as the intervention component.

However, such a conventional model has to estimate many hyper-parameters because it needs to estimate the "true" population y_t in the structural time series model equation (2.2). Basically the differences between the model predicted value and observed value provide the source for measuring the statistical impact. Such relatively complicated model identification and prediction can be vulnerable to a rapid real world change for which the model may not be able to account for during the parallel run period. If our sole goal is to estimate the statistical impact rather than produce a "true" population estimate, the model can be simplified for a parallel run scenario.

In the case of the LFS, the existing composite estimator will continue to be used to produce the "true" labour force population estimates As such the conventional intervention analysis can be simplified by modelling the differences between the estimates produced under the current design and the estimates produced under a new design conducted in parallel. This will reduce the risks of rapid changes or outliers impacting our estimation and improve robustness by reducing the model complexity. We develop the difference SSM for estimating statistical impact below.

Conceptual decomposition of a statistical impact on the GREG estimate

Suppose a new LFS design (n) starts from time t_1. Then for $t \geq t_1$ the new GREG estimate for wave i is $\hat{y}_{i,t}^{(n)}$

$$\begin{pmatrix} \hat{y}_{1,t}^{(n)} \\ \vdots \\ \hat{y}_{8,t}^{(n)} \end{pmatrix} = \begin{pmatrix} y_t \\ \vdots \\ y_t \end{pmatrix} + \begin{pmatrix} b_1^{(n)} \\ \vdots \\ b_8^{(n)} \end{pmatrix} + \begin{pmatrix} e_{1,t}^{(n)} \\ \vdots \\ e_{8,t}^{(n)} \end{pmatrix} \tag{3.1}$$

where $b_i^{(n)}$ and $e_{i,t}^{(n)}$ are RGB and sampling error of the new LFS design.

The statistical impact for each wave is

$$\underbrace{\begin{pmatrix} \hat{y}_{1,t}^{(n)} - \hat{y}_{1,t} \\ \vdots \\ \hat{y}_{8,t}^{(n)} - \hat{y}_{8,t} \end{pmatrix}}_{\text{difference in estimates}} = \underbrace{\begin{pmatrix} b_1^{(n)} - b_1 \\ \vdots \\ b_8^{(n)} - b_8 \end{pmatrix}}_{\text{difference in RGB}} + \underbrace{\begin{pmatrix} e_{1,t}^{(n)} - e_{1,t} \\ \vdots \\ e_{8,t}^{(n)} - e_{8,t} \end{pmatrix}}_{\text{difference in SE}} \tag{3.2}$$

The structural changes come from

1. a permanent level shift (LS) presented in the "difference in RGB" and
2. a dynamic sampling error change presented in the "difference in SE".

The "true" population y_t cancels out under the difference model formulation and is, therefore, excluded from estimation.

Estimating the statistical impact during a parallel run

In practice, a new design will usually be introduced by each successive rotation group. Assuming a parallel run is conducted for $t_0 \leq t < t_1$ a new series $\hat{y}_t^{(\tau)}$ can be constructed[8] as

$$\hat{y}_{i,t}^{(\tau)} = \begin{cases} \hat{y}_{i,t}^{(n)} & t_0 \leq t < t_1 \text{ and wave } i \text{ has a treatment sample} \\ \hat{y}_{i,t} & \text{otherwise} \end{cases}, \quad i = 1, \ldots, 8 \tag{3.3}$$

$$\hat{y}_{i,t}^{(\tau)} = y_t + b_i + x_{i,t}\left(b_i^{(n)} - b_i\right) + e_{i,t} + x_{i,t}\left(e_{i,t}^{(n)} - e_{i,t}\right) \tag{3.4}$$

with an intervention dummy variable $x_{i,t}$

$$x_{i,t} = \begin{cases} 1 & t_0 \leq t < t_1 \text{ and wave } i \text{ has a treatment sample} \\ 0 & \text{otherwise} \end{cases} \tag{3.5}$$

8 Note: the sampling error for the treatment sample may be larger due to smaller sample size of the treatment sample.

Therefore, the permanent level shift (LS) $\alpha_i = \left(b_i^{(n)} - b_i\right)$ (or $b_i^{(n)} = b_i + \alpha_i$ in model (2.1)) for wave i can be estimated from the parallel run with a combined sampling error process $\eta_{i,t} = e_{i,t}^{(n)} - e_{i,t}$

$$\hat{y}_{i,t}^{(\tau)} - \hat{y}_{i,t} = \alpha_i x_{i,t} + \eta_{i,t} x_{i,t} \quad t_0 \le t < t_1 \tag{3.6}$$

Assuming the sample rotation design continues, both $e_{i,t}$ and $e_{i,t}^{(n)}$ follow the same AR (see eqns (2.3) – (2.5)) process but with a different disturbance variance $\sigma^2_{\eta,\delta}$. ie.

$$\eta_{i,t} = \phi_1 \eta_{i,t-1} + \phi_2 \eta_{i,t-2} + \delta_{i,t}, \qquad \delta_{i,t} \cong NID(0, \sigma^2_{i,\delta}) \tag{3.7}$$

$$\sigma^2_{i,\delta} = \sigma^2_{i,(e,u)} + \sigma^2_{i,(e^{(n)},u)} - 2corr(e^{(n)}_{\cdot,\cdot}, e_{\cdot,\cdot})\sigma_{i,(e,u)}\sigma_{i,(e^{(n)},u)} \tag{3.8}$$

ϕ_1, ϕ_2 and $\sigma^2_{i,(e,u)}$ can be estimated from the existing LFS sample design. $\sigma^2_{i,(e^{(n)},u)}$ can be determined by the new treatment sample design. A more accurate estimate of α_i from equations (3.6) and (3.7) can be achieved by maximising correlation $\rho = corr\left(e^{(n)}_{\cdot,\cdot}, e_{\cdot,\cdot}\right)$ in equation (3.8). This relies on the working assumptions made earlier that the existing and new LFS designs have the same sampling error stochastic process, i.e. the same autoregressive coefficients of the AR(2) model. Therefore, $\sigma^2_{i,\delta} \approx (\sigma_{i,(e^{(n)},u)} - \sigma_{i,(e,u)})^2$ when $\rho \approx 1$.

The difference between the rotation group bias of the existing design and new design can be estimated from the following state space model presentation.

Observation equation:

$$\hat{y}_{i,t}^{(\tau)} - \hat{y}_{i,t} = \alpha_i x_{i,t} + \eta_{i,t} x_{i,t} \qquad i = 1, \cdots, 8 \tag{3.9}$$

State equation[9]:

$$\begin{pmatrix} \boldsymbol{\eta}_{\mathbf{t}} \\ \boldsymbol{\eta}_{\mathbf{t}-1} \end{pmatrix} = \begin{pmatrix} \Theta_1 & \Theta_2 \\ \mathbf{I} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_{\mathbf{t}-1} \\ \boldsymbol{\eta}_{\mathbf{t}-2} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\delta}_{\mathbf{t}} \\ \mathbf{0} \end{pmatrix} \tag{3.10}$$

[9] Without losing generality, the state equation is written as AR2 process although wave 1 and 2 are following a white noise and an AR1 process respectively.

Where $\boldsymbol{\eta}_\mathbf{t} = \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \\ \vdots \\ \eta_{7,t} \\ \eta_{8,t} \end{pmatrix}$, $\boldsymbol{\delta}_\mathbf{t} = \begin{pmatrix} \delta_{1,t} \\ \delta_{2,t} \\ \delta_{3,t} \\ \vdots \\ \delta_{7,t} \\ \delta_{8,t} \end{pmatrix}$

$$\boldsymbol{\Theta}_1 = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \phi_1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \phi_1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & \phi_1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \phi_1 & 0 \end{pmatrix}, \quad \boldsymbol{\Theta}_2 = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \phi_2 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & \phi_2 & 0 & 0 \end{pmatrix},$$

Analytical solution for the parallel run parameters

The estimated coefficients $\hat{\alpha}_i \ (i = 1, \cdots, 8)$ are the permanent level shift and can be derived from the new LFS design RGB as $\hat{b}_i^{(n)} = \hat{b}_i + \hat{\alpha}_i \quad (i = 1, \cdots, 8)$ where $\hat{b}_i$ is estimated from equation (3.9) and (3.10) for the existing LFS design.

The null hypothesis for no statistical impact is H_0: $\alpha_i = 0 \ (i = 1, \cdots, 8)$. Based on the classical statistical theory, we wish to determine the sample size needed to test whether the mean of the treatment samples is different to that of the control samples where the control is regarded as the true value and the difference is the statistical impact.

Temporarily assuming the term $\eta_{i,t} x_{i,t}$ in equation (3.9) is an *NID* sampling error with standard deviation $\sigma_{i,\eta} = \sigma_e$ and without considering the "systematic" sampling error process, we can derive the treatment sample sizes from equation (A1.1) and (A1.2) in Appendix 1 with $u_0 = 0$ for a single month.

For the ABS LFS, the population standard deviation σ can be roughly estimated from the sample standard error, σ_e ie, $\sigma = \sigma_e \sqrt{n_C}$ where n_C is the sample size for the control sample of 30000 households per month on average. However, e_t is actually driven by the sample error disturbance, u_t, through an AR process (see eqns (2.3) – (2.5)). In other words, the sampling error process is partly "predictable" from the AR models driven by the sampling error disturbances.

The variance of the sampling error disturbance $\sigma_{i,\mu}^2$ can be rewritten in equations (3.7) and (3.8) as

$$\begin{aligned}
\sigma^2_{i,\delta} &= \sigma^2_{i,(e,u)} + \sigma^2_{i,(e^{(n)},u)} - 2\rho\sigma_{i,(e,u)}\sigma_{i,(e^{(n)},u)} \\
&= \gamma_i\left(\sigma^2_{i,e} + \sigma^2_{i,e^{(n)}} - 2\rho\sigma_{i,e}\sigma_{i,e^{(n)}}\right) \\
&= \gamma_i\left(\frac{1}{n_C} + \frac{1}{n_T} - \frac{2\rho}{\sqrt{n_C n_T}}\right)\sigma^2_{i,e} \\
&= \gamma_i\left(1 + \frac{1}{\kappa} - \frac{2\rho}{\sqrt{\kappa}}\right)\left(\frac{\sigma_{i,e}}{\sqrt{n_C}}\right)^2 \\
&= \gamma_i\left(1 + \frac{1}{\kappa} - \frac{2\rho}{\sqrt{\kappa}}\right)\sigma^2_{i,e}
\end{aligned} \tag{3.11}$$

where $\sigma^2_{i,e}$ and $\sigma^2_{i,e^{(n)}}$ are the variance of the control and treatment sampling errors of wave i respectively.

$$\gamma_i = \begin{cases} 1, & i = 1\ (\phi_1 = 0, \phi_2 = 0) \\ 1 - \phi_1^2, & i = 2\ (\phi_1 \neq 0, \phi_2 = 0) \\ (1+\phi_2)[(1-\phi_2)^2 - \phi_1^2]/(1-\phi_2), & i \geq 3\ (\phi_1 \neq 0, \phi_2 \neq 0) \end{cases}$$

is derived from sampling error AR process.

$\kappa = n_T/n_C$ is the ratio between treatment and control sample sizes.

Considering the problem at hand is one sample of treatment samples, the variance of treatment samples can be derived as

$$(n_C/n_T)\sigma_e^2 = \sigma_e^2/\kappa\,.$$

Replacing $\sigma/\sqrt{n}$ in equations (A1.1) and (A1.2) by the standard deviation of sampling error of treatment samples, $\sigma_e/\sqrt{\kappa}$, we produce equation (3.12) to determine the treatment sample size n_T.

$$n_T = \left(\sigma_e\sqrt{n_C}\,\frac{z_{1-\alpha/2} + z_{1-\beta}}{a_i}\right)^2 \text{ where } z = \frac{\alpha_i}{\sigma_e/\sqrt{\kappa}} \tag{3.12}$$

After considering the AR2 sampling error process and intra cluster correlation from (3.11), we derive the standard error of α_i at any point of time t

$$SE(\alpha_i \mid \kappa) = SE(\delta_{i,t}) = \sigma_{i,\delta} = \sqrt{\gamma_i\left(\frac{1}{\kappa} + 1 - \frac{2\rho}{\sqrt{\kappa}}\right)}\sigma_{i,e} \tag{3.13}$$

We can also derive that the improvement is the gain

$$SE(\alpha_i \mid \kappa)\Big/\frac{\sigma_{i,e}}{\sqrt{\kappa}} = \sqrt{\gamma_i\left(1 + \kappa - 2\rho\sqrt{\kappa}\right)} \tag{3.14}$$

in terms of the proportional reduction to the standard error of α_i by considering sampling error process, and intra cluster correlation using the SSM model. The smaller the gain value implies the bigger the reduction of the standard error. This gain decreases with increasing intra cluster correlation ρ. When $\rho = 0$ (there is no intra cluster correlation), the gain is $\sqrt{\gamma(1+\kappa)}$. For example, when $\kappa = 0.5$, then the gains are 0.64 and 0.96 for employed and unemployed of the LFS respectively.

For a parallel run with a treatment sample over periods $\{T\}$, the standard error of α_i is

$$SE(\alpha_i \mid \kappa, n) = \sqrt{\frac{1}{n}\sigma^2_{i,\delta}} = \sqrt{\frac{\gamma_i}{n}\left\{\frac{1}{\kappa} + 1 - \frac{2\rho}{\sqrt{\kappa}}\right\}}\sigma_{i,e} \text{ because } \delta_{i,t} \cong NID(0, \sigma^2_{i,\delta}) \qquad (3.15)$$

where n is the number of times of wave i is observed over the periods of $\{T\}$.

3.2 Simulation Study

Equation (3.13) provides a theoretical solution to determine the standard error of the statistical impact $\{\alpha_i\}$. It can be used to allocate the treatment sample size by optimising n (the number of times each wave is included in the periods of a parallel run) and κ (treatment sample size proportion to control sample size) to meet the statistical accuracy criteria with the predefined parameters γ_i, ρ and $\sigma_{i,e}$ (which are specific for employed and unemployed estimates). For this simulation study we assume an equal sampling error for the 8 waves[10], i.e. $\sigma^2_{i,e} = \sigma^2_e \ (i = 1, \cdots, 8)$. The sampling error disturbance variance of wave *i,* $\sigma^2_{\delta,i}$ can be derived from

$$\sigma^2_{i,\delta} = \gamma_i \left(1 + \frac{1}{\kappa} - \frac{2\rho}{\sqrt{\kappa}}\right)\sigma^2_e$$

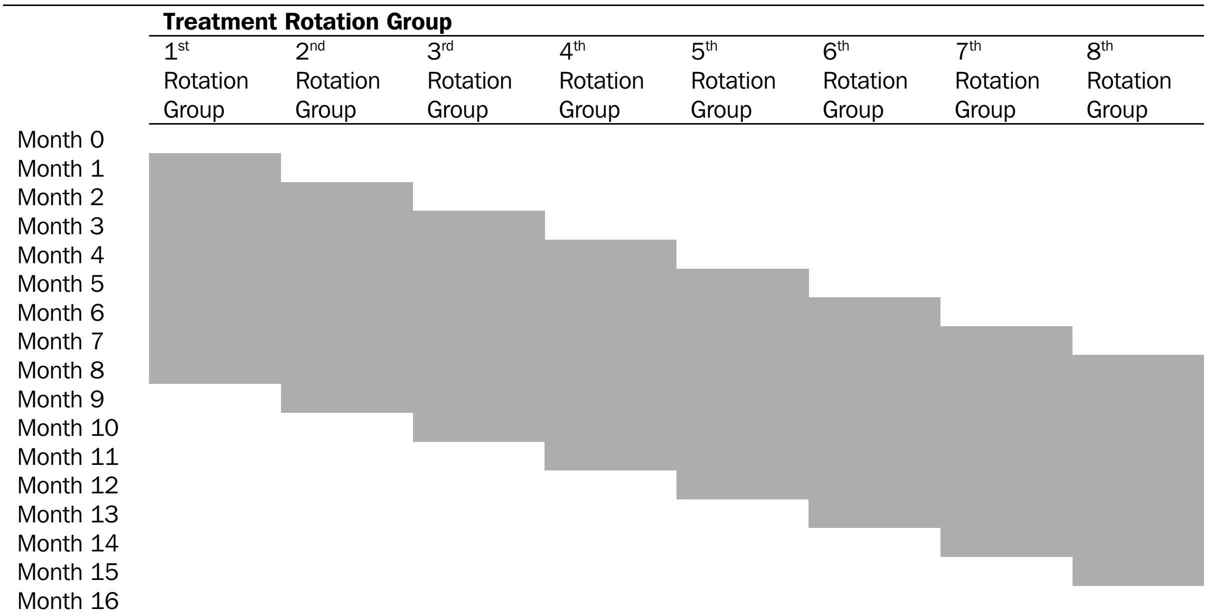
A simulation study was undertaken with the following two objectives:

- to verify whether the theoretical solution is correct, and
- to evaluate Kalman filter performance of the SSM on relatively short time series derived from a parallel run.

Due to operational constraints, only one treatment RG can be introduced per month. Figure 3.1 presents graphically a 15 month parallel run scheme, with each treatment rotation group running in parallel for a full 8 (=15-7) months. The shaded cells represent treatment rotation groups.

[10] An alternative assumption is the equal sampling error disturbance variances. We found the two assumptions do not make difference for the standard error of α_i because they have the same internal sampling error structure driven by AR2 process.

3.1 Scheme for 15 months parallel run



100 replicates were generated for different combinations of

- parallel durations (11, 13, 15, and 19 months),
- treatment sample sizes (κ = 30%, 50%, 80% and 100%), and
- intra cluster correlations (ρ =0, 0.3, 0.5, 0.8)

Appendix 3 gives more details on how the data were generated for this simulation study.

Zero means and the known sampling errors variances for both control and treatment samples are used to set up the initial condition for sampling error disturbances $\boldsymbol{\mu}_0$ for the Kalman filter to start with. See more details in Section 4.

The standard errors of $\{\alpha_i\}$ were estimated and found consistent over different waves (i=1,2,…8) regardless of the true value of $\{\alpha_i\}$. The top panel of Figure 3.2 shows the average standard error of α_i (*avg.se*[11]) against different combinations of intra cluster correlation (Rho), parallel run duration (Period) and treatment sample size (Kappa). The results are consistent with our expectation for both employed and unemployed. That is, the larger intra cluster correlation or the longer parallel run duration or larger the treatment sample size, the smaller of the standard error of α_i.

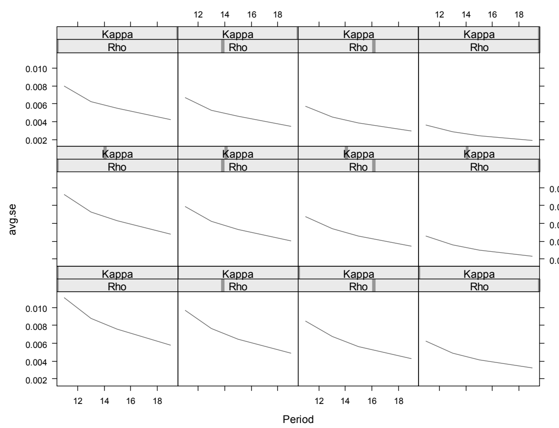
[11] The estimated standard error of α_i for each replicate appears consistent regardless of the waves and the size of α_i. *avg.se* is calculated as the average of all replicates across 8 waves ie.

$$avg.se = \frac{1}{8\times 100}\sum_{i=1}^{8}\sum_{j=1}^{100}\hat{\sigma}_{i,j} \text{ where } \hat{\sigma}_{i,j} \text{ is the estimated standard error of } \alpha_i \text{ (wave } i\text{) for replicate } j.$$

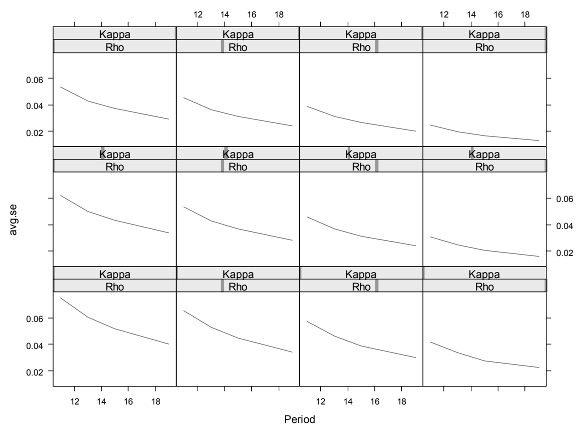
We also created a variable $X = \sqrt{\frac{1}{n}\left(1 + \frac{1}{\kappa} - \frac{2\rho}{\sqrt{\kappa}}\right)}$ to examine its relation with the standard error of α_i. The lower panel of Figure 3.2 shows the simulated results (dots) against the result (regression line) derived from regressing *avg.se* on to *X*. This graphical presentation clearly demonstrates that there is a very strong linear relationship between *avg.se* and *X*.

3.2 Average SE against treatment sample size, intra cluster correlation, and parallel duration

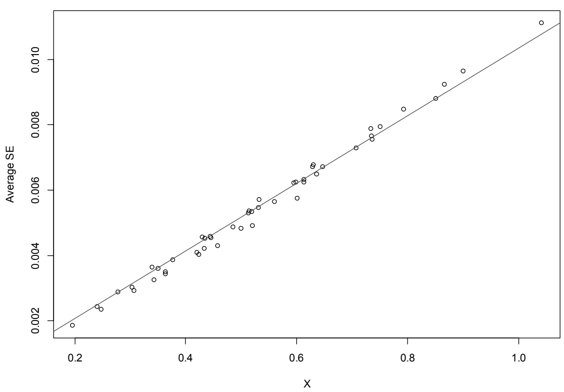
Average SE of estimated SIM of the Simulated Employed Data



Average SE of estimated SIM of the Simulated Unemployed Data



Average SE against X of the Simulated Employed data



Average SE against X of the Simulated Unemployed data

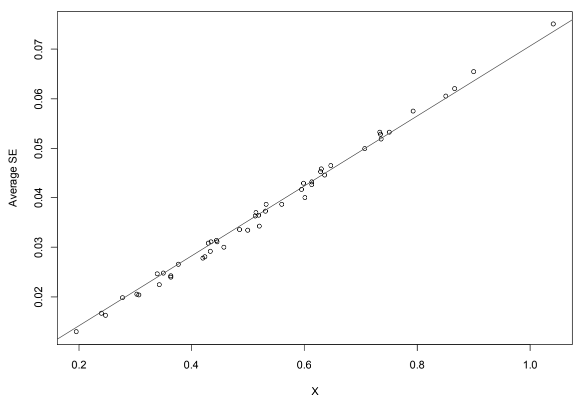


Table 3.3 shows the regression analysis result. The high R^2 values confirm that *avg.se* can be predicted from *X*.

From this analysis, we can confidently conclude that the theoretical articulation of equation (3.13) is correct. It appears that the Kalman filter performed well to estimate α_i with expected standard error. More detailed analysis on Kalman performances will be discussed in the next section.

3.3 Average SE Regress on X

	Employed	Unemployed
Coefficient of X	1.035e-02 (5.482e-05)	0.0706435 (0.0002898)
Null deviance	1.6335e-03	7.6127e-02
Residual deviance	2.1526e-06	6.0167e-05
R^2	0.9987	0.9992

In the context of the parallel run design, from the structure of *X*, we confirm that

- Intra cluster correlation is the most powerful variable to reduce the standard error of estimated α_i.
- The treatment sample size is the second most important variable. When the treatment sample size is the same as the control sample size (ie. $\kappa = 1$), it is the most efficient balanced design to minimise the standard error of estimated α_i.
- The duration of parallel run is the least powerful factor among the three to reduce the standard error of estimated α_i.

The coefficients of *X* in Table 3.3 can be used with any combination of parallel run duration, treatment sample size and intra cluster correlation to predict the standard error of α_i for both employed and unemployed of the LFS. Therefore, an optimised parallel run design can be achieved.

However, the intra cluster correlation between control and treatment samples is usually unknown[12], unlike the sampling error and the rotation panel design induced AR sampling error dynamics which can be estimated from the sample data (see Pfeffermann et al., 1998).

Using LFS historical data, samples in each rotation group were split into pseudo control and treatment groups, and then used to estimate their intra cluster correlation. Appendix 2 presents some details of this study. We concluded that the intra cluster correlation between the control and treatment groups was reasonably high for employed and low for unemployed. This result is within our expectation because employed persons are likely to keep their status in the two groups, therefore, resulting in a high intra cluster correlation. In contrast, unemployed persons change their status more often in the two groups, therefore their intra cluster correlation is likely to be low. Our study demonstrated the existence of the intra cluster correlation within the same sample design of the LFS.

The future ABS LFS will be based on an address register list based master sample rather than the current and traditional area master sample. Therefore, the intra cluster correlation between the control and treatment samples will not be known. However, this study suggests that we should explore possibilities to increase the intra cluster

[12] The intra correlation between control and treatment samples can be established if a block experiment design is implemented for both control and treatment samples. See the detailed discussion in Section 4.

correlation between control and treatment groups for any future parallel run. One such idea is to design the parallel run as a randomized block design, using the first phase geographical areas as block variables. The details will be discussed in the next section.

How the wave level statistical impact $\{\alpha_i\}$ affect the overall statistical impact to the composite estimate is articulated in next section and appendix 4 in relation to the correlations among the estimated $\{\alpha_i\}$.

4. MEASURING STATISTICAL IMPACTS WITH A SHORT TIME SERIES FROM A PARALLEL RUN

4.1 Issues of Kalman filter initialisation and convergence

When using an SSM to estimate state variables, the Kalman filter is a commonly used tool. It is a recursive algorithm producing optimal results in terms of mean square error when being applied to linear models. The Kalman filter produces filtered estimates of state variables at time t by taking the mean and variance of that state conditional on observations up to time t-1 as an input. Therefore, initialisation at the start point t=0 is a necessary computational step and such initialisation may be accomplished in a variety approaches. If no prior information is available, a diffuse initialisation of the Kalman filter is commonly used for non-stationary state variables, which implies that the initial values have zero mean and very large variance. As a result, state estimation will have a transient effect in the initial phase of Kalman filtering before it converges to its steady state.

With a diffuse prior, the Kalman filter estimates all the state variables, and their covariance matrices by using the parallel run data, even though some of the related information is readily available. While this approach can be useful for approximate exploratory work, it is not recommended for general use because it is not efficient and can lead to large error when the time series is short. A more efficient approach is to use available a-priori information in an exact initialization of the Kalman filter. Therefore, Kalman filter can converge more quickly and produce the state variable of interests with better quality and suffers less of the side effect from insufficient convergence.

Faster convergence enables us to achieve accurate state estimates in a shorter parallel run, which is desirable from a cost and operational feasibility point of view. Because the SIM model is stationary with stability and observability, therefore, the unconditional mean and covariance can be used as the initial condition of the state variables (Aoki 1987). We set the initial conditions for

(1) the state $\{\alpha_i\}$ with values of zero and large variances[13] with a known correlation structure.(see details in Appendix 4),

(2) the sampling error state $\{e_i\}$ with zero expectations, estimated variance and zero correlations.

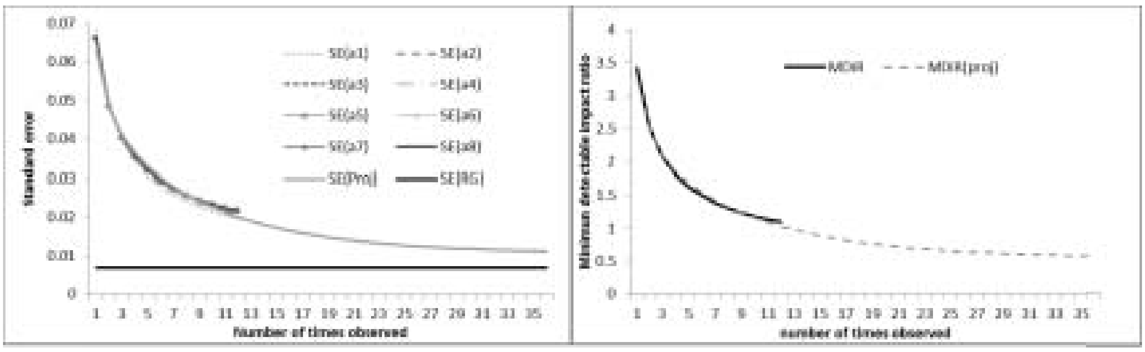
The covariance of the Kalman filter estimated $\{\alpha_i\}$ is the state covariance matrix which has a special form because of the special SSM structure. In appendix 4, we derive the analytical solution of the corresponding state correlation matrix which is the sum of the serial cross-correlation of sampling errors $\{e_i\}$.

Subsequently, the standard error of the statistical impact to the composite estimate can be approximated by applying a multiplier to the standard error at wave level. 0.6798 and 0.8681 are the multiplier values for LFS unemployed and employed respectively (see the derivation details in Appendix 4).

We evaluated whether the Kalman filter worked effectively for the SSM by simulating 100 replicates based on the feature of unemployed of LFS with $\kappa = 1$, $\rho = 0$ and each wave had 12 observations, ie. the treatment group had the same sample size as the control group; there was no intra cluster correlation between them; the duration of the parallel run was 19 months.

Figure 4.1 shows how the SSM performed. The left panel plots the average standard error of filtered estimates of impact $\{\alpha_i\}$ against the number of observed periods n of parallel data collection (= duration of parallel run minus 7) as the horizontal axis. SE(Proj) is the projected standard error[14] beyond 12 observation periods. SE(RG) is the reference for the theoretical standard error of each wave when they reach their steady states.

4.1 Standard error of filtered estimate of statistical impact at RG level for for unemployment



---

[13] A priori information should be used if available. As mentioned in Section 1, experimental design and field tests can be utilised to obtain such priori information. How it can be done is out scope of this paper.

[14] The projected standard error is produced by modelling empirically the simulated results up to 12 observation periods. It illustrates the trajectory of the Kalman filter convergence property.

From the left panel of Figure 4.1, it is clear that the standard errors of $\{\alpha_i\}$ are consistent regardless of the true values of $\{\alpha_i\}$. The projected standard error (SE(RG)) demonstrates that the standard error of the filtered estimates will continue to converge to the theoretical value by increasing the number of observations over time.

The question is whether the filtered estimate is accurate enough to meet the predefined accuracy criterion? (See the details in Section 3). To answer this question, we calculated the equivalent MDI ratios (MDIR) for the national level composite estimate and plotted in the right panel of Figure 4.1. The MDIR is 1.09 which is slightly short for reaching the predefined accuracy criterion after 12 observation periods, ie. 19 months parallel run duration.

It appears that under this approach 13 observations (i.e. parallel run length of 20-months) is required to detect one standard error impact for the national level composite estimate. Given that the Kalman filter approach is subject to a transient period, we need to look into how to improve its convergence rate. This can be achieved by providing better prior information for Kalman filter initialisation, for example the Hybrid Option in section 5 and other statistical methods from randomized experiments as a part of SIM, that is, to find more efficient designs for the parallel run.

4.2 Sample design based experimental approach

In Section 3, an SSM for the parallel run (eqn. 3.9 - 3.10) is proposed. The improvement in precision due to the positive correlation between survey errors of the regular survey and the treatment survey comes from the intra-cluster correlation, since samples are drawn from the same primary sampling units. Taking advantage of this intra-cluster correlation in our power calculations implies that the parallel run to measure a level shift $\{\alpha_i\}$ must be designed as a Randomized Block Design (RBD) with the cross-classification of Primary Sample Units (PSU) and months as the block variables in the experimental design. By doing so, the variation between blocks can be eliminated from the variance of the statistical impact estimate. Van den Brakel and Renssen (2005) have studied how to design embedded experiments to detect and quantify possible level shifts in time series due to necessary changes to a sample survey, providing a safe transition from an old to a new survey design. They developed design-based procedures for the analysis of large scale experiments embedded in sample surveys, for example to estimate the effect of a redesign on the outcomes of a sample survey. Variance estimates derived in their paper directly utilise the gain in precision based on the fact that observations within blocks (PSU times month) are more homogeneous. Instead of estimating the variance of a contrast as the variance of the sum of the sampling error of the regular and treatment survey minus (two times) the correlation between both errors, they directly estimate the net variance by removing the variation between blocks from the estimated treatment effect.

This method has several advantages:

1. The analysis is based on the micro-level, therefore has more degrees of freedom for variance estimation, and avoids the problem of short time series like those used in the SSM.
2. The method is design-based which facilitates the interpretation of the results. This means that the method tests hypotheses about differences between survey estimates observed under two different approaches and does not assume some kind of model. An additional advantage of this is that the estimated statistical impacts are consistent with domains.

The SIM phase 1 (described in Section 1) would also benefit from such an experimental design to meet the requirement of testing numbers of changes separately in the SIM phase 1. Its results can then be utilised as inputs to the SSM model for the parallel run.

Further study is needed to understand the merit of this approach, how it relates to the SSM approach, and to evaluate its feasibility for field operations and power.

5. DIFFERENT OPTIONS FOR MEASURING STATISTICAL IMPACT AND CHANGE IMPLEMENTATION

The following types of options have been considered by the ABS as a part of this study.

Option A: 100% control sample to maintain the current LFS production quality during the parallel run, while finding optimal combinations of the treatment sample size and length of parallel run. This option has the lowest risk level but would be costly.

Option B: Reduce the size of the control sample and make the treatment sample size equal to the control sample size, e.g. a control and treatment group equal to 75% of the regular sample size of the current LFS. The aim of this balanced design is to estimate the statistical impact as precisely as possible at the cost of accepting more volatile regular LFS estimates for official publication purposes during the parallel run period. This option might not be accepted by external users due to increase of sampling error in regular survey estimates.

Option C: Phase-in the new process with one rotation group each month at a time. After 8 months, the existing process is fully changed-over to the new process. This strategy does not have a period of parallel data collection, thus statistical impact measurement fully relies on a time series model (see eqn. 2.1 – 2.5) to estimate the statistical impact. A potentially large revision may result and have to be accepted after starting the changeover. This is considered to be the highest risk option.

The remainder of this section assesses the three options and proposes a hybrid option if an additional revision is acceptable 12 months after the introduction of a new LFS survey.

5.1 Evaluation of the three options

From the formulae developed in Section 3, we can calculate the parallel run parameters based on a given set of scenarios. We assumed the intra cluster correlation between control and treatment samples is zero (i.e. $\rho = 0$) because a new LFS design is hypothetical at this stage. Table 5.1 shows the length of the parallel run required for scenarios of Option A with a 100% and 50% treatment sample (A100 and A50, respectively), and Option B with 75% (B75) for both control and treatment samples to meet the predefined accuracy criterion (see Section 3). It appears that none of the options can meet the defined accuracy criterition with the operational feasibility constrains of a parallel run shorter than 24 months.

5.1 Sample size and length of parallel run required for Option A and B

	A100	A50	B75
Control sample size %	100	100	75
Treatment sample size %	100	50	75
SE on published estimates	Current	Current	1.15 times larger
Duration	32 months	>44	36 months
Risk	Low	Low	Moderate

Options A50 and B75 have the same total samples per month and their costs should be similar. However, B75 is a balanced design and is more efficient than A50 for measuring statistical impact. Therefore, a shorter parallel run is sufficient, at the cost of more (1.15 times) volatile published LFS estimates during parallel run periods because of sample size reduction for control samples.

For the unemployment example with a 24 months parallel run, the estimated relative standard errors of a statistical impact for the composite estimates are 1.65%, 2.02% and 1.91% from the three options respectively. They detect one standard error (2.6%) statistical impact with 5% type I error, and powers of 47%, 36%, and 39% respectively. Another interpretation is that the three options can detect the size of statistical impact of MDI ratios 1.24, 1.52 and 1.43 times the current survey standard error (2.6%) with the pre-defined precision.

Phase-in (Option C) is an implementation strategy without parallel data collection and is not designed for accurate statistical impact measurement. This approach is of a high risk since there is limited opportunity to appraise the impacts before implementation.

A one standard error statistical impact is not detectable with the required accuracy within the 8 months phase-in period because the statistical impact is wave sensitive and there are incomplete or insufficient observations of new samples. Table 5.2 illustrates our simulation results with a one standard error statistical impact for LFS unemployed (2.6%), and the impact is measured at 3, 5 and 8 months for 100 replicates with an unrealistic assumption that the impact is uniform to all the waves (wave insensitive impact).

5.2 Total level shift detected by SSM across 100 replicates (Unemployed)

Periods after the first new design rotation group is introduced (month)	Overall impact %	MDI ratio
Simulated	2.60	
3	1.86	4.1
5	2.23	4.0
8	2.34	3.9

The measured impacts at 3 and 5 months (1.856% and 2.232% respectively) are obviously not accurate. The MDI ratio indicates that an impact of more than 3.9 standard error of unemployed population estimate can be detected with the required precision. Under this circumstance, there are two choices of how to handle the situation if this option was applied alone:

1. Ignore the impact because the measured impact cannot meet the accuracy criterion. The statistical impact will be in the published estimates and could be misinterpreted as real world change.

2. Apply an adjustment based on the estimated impact. This action is ad-hoc and potentially subject to large revision later.

Neither of these choices is considered acceptable, therefore, this is a very high risk option.

Table 5.2 also shows that the estimated impact is close to the real impact after the end of phase-in (8 months). However, the impact cannot be measured accurately even after 24 months. The details are presented in the next sub-section.

5.2 Simulation Studies for Different Options and a Hybrid Option

The advantage of a large parallel run (Option A) is that it minimises the risk for regular publication purposes during the changeover. If unexpected results are observed with the new process during the parallel run there is still the flexibility to fall back on the old approach. Because a large parallel run can estimate the statistical impact directly with high precision, another advantage is that it facilitates the implementation of the new survey without further revision of the impact measurements after the changeover. This approach, however, is not cost effective since a significant data collection effort is required.

The opposite approach (Option C) is having no parallel run and estimating the impact using a time series model. The major advantage of this approach is that it is inexpensive and avoids the additional fieldwork required for a parallel run. Skipping a period of parallel data collection and relying on a time series model to estimate the statistical impact also has several disadvantages and risks. First, it is not clear in advance if the time series estimates for the statistical impact will reach the required precision. Furthermore, estimates for the impact are unreliable directly after the changeover, and are likely to be revised after new observations become available under the new survey design. As a consequence, revisions must be expected and accepted. Implementing the changeover without a period of parallel data collection increases risk during the changeover. If the new survey design turns out to be a failure or has a significant impact, and it is therefore decided to fall back on the old approach, then there is a period where no data or less reliable data are available for the production of official statistics.

An intermediate option is to have a small parallel run and combine this information with a time series modelling approach. We start for example with a parallel run with 20% or 50% of the regular sample size for a period of 12 months. The results obtained from the parallel run could be used as a-priori information in the time series model. The information that becomes available under the new approach would be used in the time series model to further improve the precision of the impact estimates. Note that this option directly reduces the risk of having a period without official figures after the changeover and also reduces the amount of revisions.

Simulations of the national unemployed and employed persons from the rotation group level estimates were conducted to illustrate the precision of the impact estimates[15]. The simulation was run with the time series model approach without a parallel run and with five different parallel run scenarios of reduced sample sizes as summarized in Table 5.3. The standard errors in Table 5.3 refer to the statistical impact estimates at the rotation group level obtained with the control sample, the treatment sample and the specified parallel run periods. The sample size percentages refer to the current sample size of the regular LFS. The Minimum Detectable Impact (MDI) ratios were calculated for the overall composite estimates.

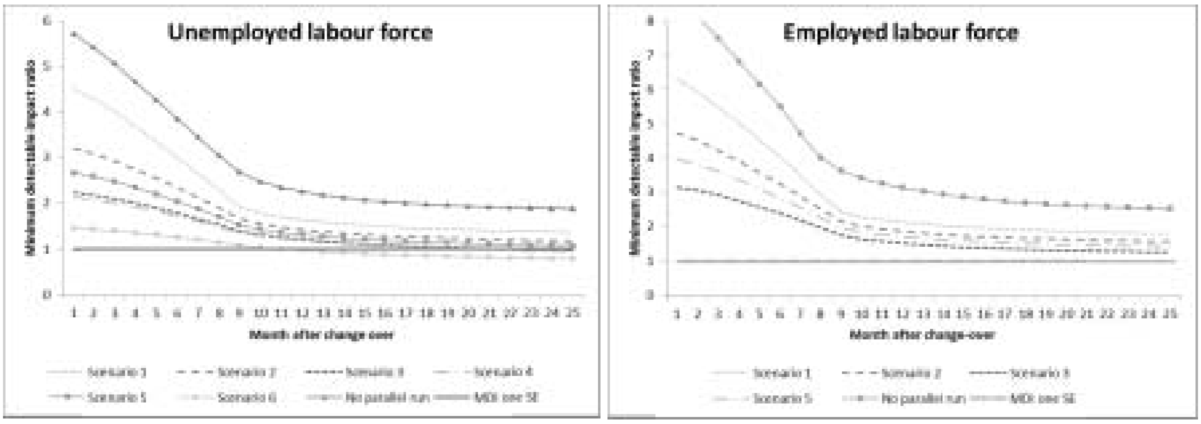
---

[15] Most precision discussions are focused on the rotation group level in this section because the standard error of a statistical impact to the overall composite estimates can be approximated as the standard error at rotation level times a multiplier.

5.3 Different scenarios of parallel run used in the simulation[16]

Scenario	Sample size		Parallel	Unemployed			Employed		
	control sample	treatment sample	run period	Standard error		MDI ratio	Standard error		MDI ratio
			month	% points	total		% points	total	
1	100%	20%	18	7.9	61620	4.05	1.08	135000	5.25
2	100%	50%	12	5.6	43680	2.87	0.81	101250	3.94
2	100%	20%	24	5.6	43680	2.87	0.81	101250	3.94
3	100%	50%	18	3.9	30420	2.00	0.54	67500	2.63
4	100%	100%	12	3.7	29016	1.90	0.54	67500	2.63
5	100%	25%	24	4.7	36348	2.41	0.68	85000	3.31
6	100%	100%	18	2.5	19500	1.28			

5.4 Minimum detectable impact ratio at 5% significance level and 50% power obtained with the time series model for different periods after the changeover for unemployed labour force (left panel) and employed labour force (right panel)



The standard errors obtained with the time series model without a parallel run and the five different scenarios are aggregated for the different periods observed after the changeover to the new design. The MDI ratio values obtained directly after the parallel run are all greater than one except for unemployed scenario six. This suggests none of the parallel run results from the first five scenarios can meet the predefined precision. Figure 5.4 depicts the MDI ratios of the different scenarios against different periods after the changeover for the unemployed and the employed. For the unemployed a sixth scenario is added to illustrate what the time series model adds if it is applied after the full parallel run of 100%-100% for a period of 18 months. The MDI ratio for the scenario without a parallel run converges to a value of about 2 and 3 for unemployed and employed respectively, which implies that under this scenario detecting an impact of one standard error cannot be achieved. For example, for the unemployed series of Scenario 1, a one standard error impact still cannot be achieved with the predefined accuracy criterion after 24 months. For Scenario 4, this precision

16 *Note that for employed scenario 4 equals scenario 3*

is obtained after 19 months, and for Scenario 6 it takes 11 months to achieve this precision.

To illustrate the volatility of the impact estimates if there is no parallel run, i.e. Option C, the left panel of Figure 5.5 shows for 10 replicates how the time series model estimates the impact if more observations become available after the changeover in a rotation group. The horizontal axis depicts the number of months observed under the new design after the changeover. As can been seen it takes about 12 months before a stable estimate for the impact in a particular wave is obtained. The right panel contains similar estimates but now combined with the information obtained with a parallel run under scenario 3. The time series model further improves the impact estimates, whilst also the volatility of the estimates directly after the changeover is clearly reduced.

5.5 Impact estimates for the unemployed for 10 replicates of one wave obtained with the time series model for different periods after the changeover without a parallel run (left panel) and with a parallel run according to Scenario 3 (right panel). The real value of the discontinuity is 15 percentage points which is equal to 117000 unemployed

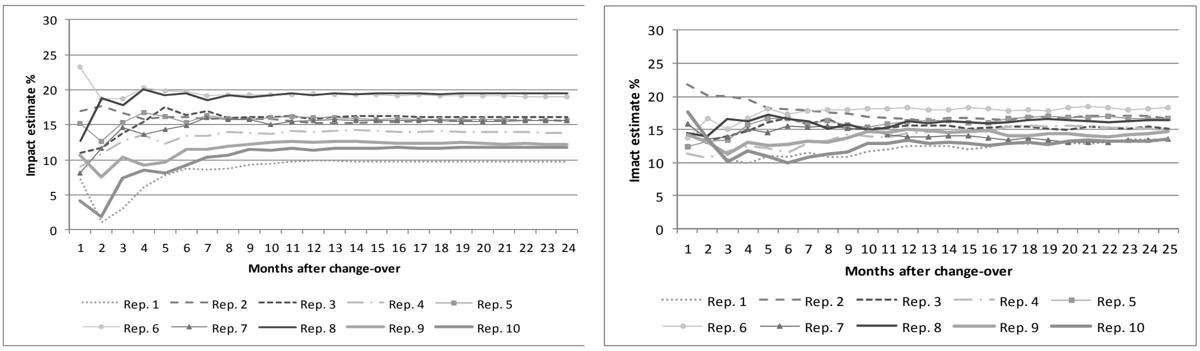
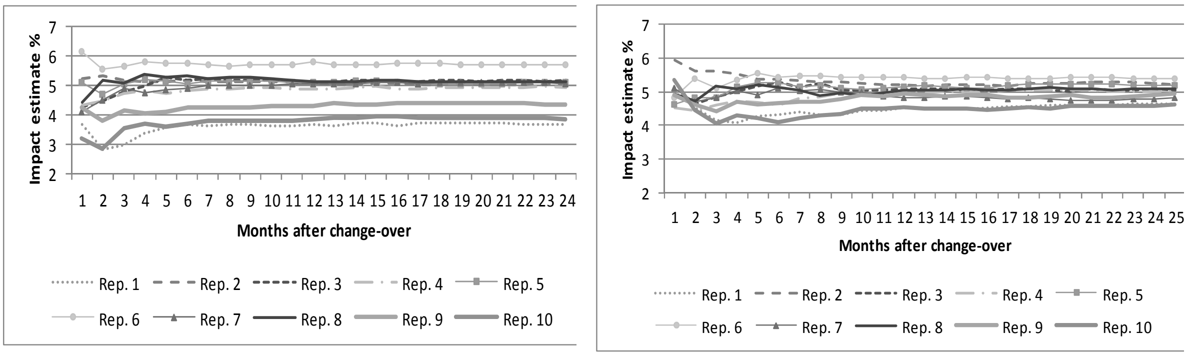


Figure 5.6 gives similar figures for the employed labour force. Estimates appear to be more stable and converge faster compared with the unemployed.

5.6 Impact estimates for the employed for 10 replicates of one wave obtained with the time series model for different periods after the changeover without a parallel run (left panel) and with a parallel run according to Scenario 3 (right panel). The real value of the discontinuity is 5 percentage points which is equal to 625000 employed.



The first part of Table 5.7 summarizes the standard errors of the impact estimates for the unemployed (in terms of percent points) in the separate waves (after 24 months). As expected the precision increases with the sample size of the parallel run. This table shows that without a parallel run, standard errors for the rotation groups are estimated with a standard error of about 3.5%. Under Scenario 1, the parallel run results in an impact estimate with a standard error of 7.9% (from Table 5.3). The time series model improves the precision to about 2.6% (last column of Table 5.7). In a similar way the standard error of a parallel run with a standard error of 5.6% is further reduced with the time series model to 2.2% (Scenario 2) and from 3.9% to 1.8% (Scenario 3).

A consequence of improving the results of a relative small parallel run with a time series model is that the initial estimates of the statistical impact obtained with the parallel run are likely to be revised after, for example, a period of 12 months. Using a simulation study, we estimated the expected amount of revisions between the estimates obtained for parallel runs under the five scenarios and the time series model after 12 months. As expected, the size of the revisions decreases with the sample size of the parallel run. The expected revision (in the "Average" column of table 5.7) is about 5.8% under Scenario 1, 4% under Scenario 2 and 2.7% under Scenario 3.

The final standard errors and revisions of the SIM estimates are given for the employed for the different scenarios in Table 5.8 in terms of percent points.

5.7 : Standard errors for impact measurement estimates and revisions for the unemployed labour force after 12 months under different parallel run options in percentage points

	RG1	RG2	RG3	RG4	RG5	RG6	RG7	RG8	Average
S.E. of impact									
No PR	3.18	3.35	3.45	3.52	3.58	3.63	3.66	3.69	3.5
Scenario 1	2.36	2.47	2.51	2.54	2.57	2.63	2.67	2.74	2.6
Scenario 2	2.05	2.13	2.15	2.17	2.19	2.23	2.29	2.37	2.2
Scenario 3	1.74	1.79	1.79	1.79	1.81	1.84	1.91	2.00	1.8
Scenario 4	1.70	1.75	1.74	1.75	1.77	1.80	1.86	1.95	1.8
Scenario 5	1.88	1.95	1.96	1.97	1.99	2.03	2.09	2.18	2.00
Scenario 6	1.42	1.44	1.43	1.43	1.44	1.47	1.52	1.61	1.47
Revision									
Scenario 1	5.62	6.13	5.88	6.53	4.71	5.71	5.54	5.93	5.8
Scenario 2	3.97	4.22	3.93	4.57	3.29	3.89	3.82	4.02	4.0
Scenario 3	2.73	2.83	2.59	3.11	2.23	2.56	2.58	2.65	2.7
Scenario 4	2.60	2.69	2.45	2.95	2.12	2.43	2.45	2.51	2.52
Scenario 5	3.29	3.46	3.17	3.77	2.71	3.15	3.13	3.25	3.24
Scenario 6	1.69	1.72	1.55	1.88	1.38	1.54	1.58	1.56	1.61

Revisions are calculated as the mean over the absolute value of the difference between the initial estimate of the parallel run and the time series estimate after 12 months after the changeover. If the size of the revisions is compared with the standard error of the SIM, it must be concluded that the revisions are still substantial, particularly in the case of small parallel runs. As expected, the size of the revision decreases as the size of the parallel run increases. As illustrated with Scenario 6 for the unemployed labour force, the time series model still produces revisions after a full parallel run designed to observe a SIM of one standard error at a 5% significance level and a power level of 50%.

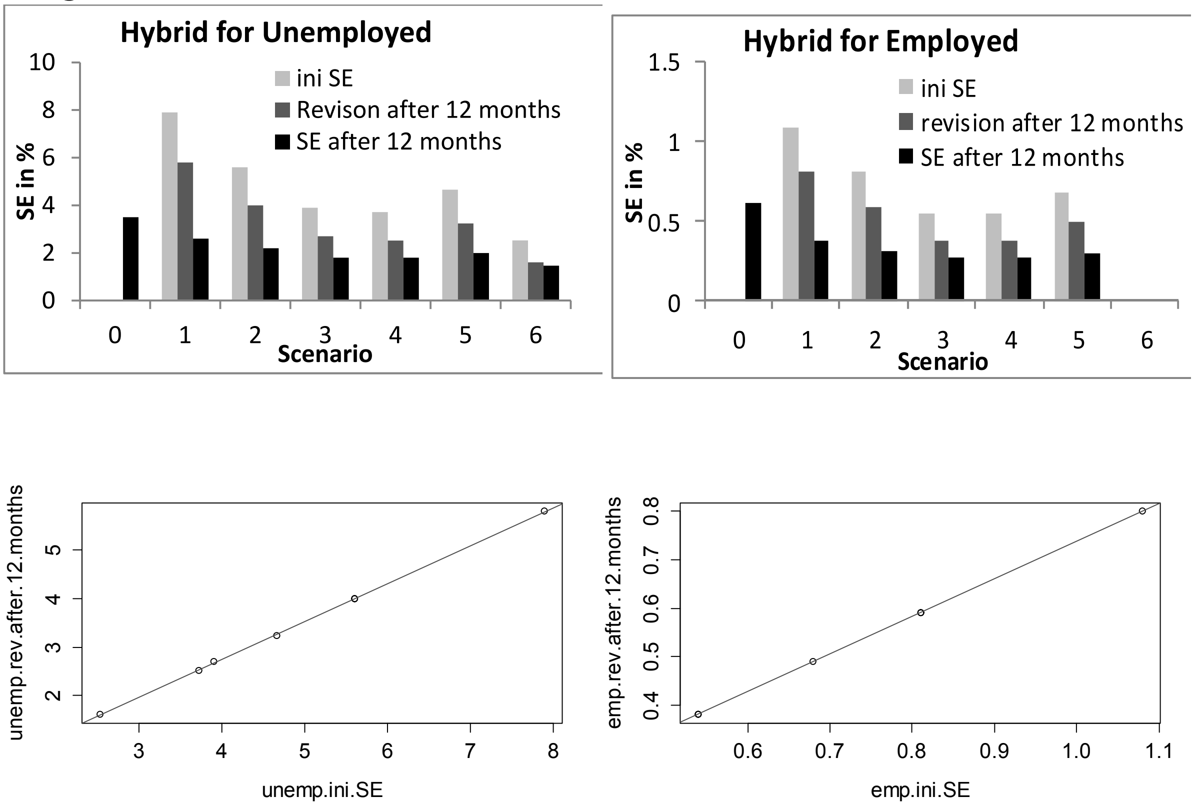
5.8 : Standard errors impact measurement estimates and revisions employed labour force in percent points after 12 months under different parallel run options

	RG1	RG2	RG3	RG4	RG5	RG6	RG7	RG8	Average
S.E. of impact									
No PR	0.57	0.58	0.58	0.61	0.64	0.63	0.63	0.65	0.61
Scenario 1	0.37	0.36	0.35	0.36	0.38	0.37	0.37	0.40	0.37
Scenario 2	0.31	0.30	0.29	0.30	0.31	0.31	0.31	0.33	0.31
Scenario 3	0.25	0.23	0.22	0.22	0.23	0.23	0.24	0.26	0.27
Scenario 4	0.25	0.23	0.22	0.22	0.23	0.23	0.24	0.26	0.27
Scenario 5	0.27	0.28	0.28	0.29	0.29	0.30	0.31	0.33	0.29
Revision									
Scenario 1	0.77	0.87	0.81	0.90	0.67	0.78	0.78	0.83	0.80
Scenario 2	0.58	0.63	0.59	0.67	0.50	0.57	0.57	0.61	0.59
Scenario 3	0.38	0.41	0.37	0.44	0.32	0.36	0.36	0.37	0.38
Scenario 4	0.38	0.41	0.37	0.44	0.32	0.36	0.36	0.37	0.38
Scenario 5	0.48	0.52	0.48	0.56	0.41	0.47	0.47	0.49	0.49

5.3 Revision analysis for the hybrid option

The purpose of this analysis is to understand the properties of the hybrid option. This option uses the initial estimates of statistical impacts (*ini SE* in light grey) from a small parallel run, which may not be as accurate as desired, as inputs to a time series model (SSM) to improve the accuracy 12 months after the changeover. In particular, the relationships between the standard error of the initial estimates (*ini SE*), the standard error of the final statistical impacts (SE 12 months after changeover) and revision size 12 months after changeover are explored.

5.9 : Comparisons of initial SE from parallel run, final SE and Revision after 12 month changeover



The top panel of Figure 5.9 plots the three sets of simulated results for both unemployed and employed under the different scenarios presented in the last sub-section. The lower panel plots the revisions in circles and the fitted value along the line by regressing the revision sizes onto the initial SEs.

It appears that the regression lines fit the simulated results very well for both employed and unemployed. Table 5.10 shows regression results and performance. Both coefficients of *ini SE* for the unemployed and employed are 0.78, and suggest the hybrid option reduces about 80% of errors regardless of the quality of *ini SE*.

5.10 Revision size regressing on ini SE

	Unemployed	**Employed**
Intercept	-0.367054 (0.021980)	-0.0395645 (0.0008405)
ini SE	0.779785 (0.004307)	0.7774410 (0.0010967)
Null deviance	10.9734000	1.2628e-01
Residual deviance	0.0016739	1.0051e-06
R^2	0.99984	0.999992

Table 5.11 below shows the results from regressing the SE of the final estimates from the hybrid option 12 months after changeover onto the initial SE (*ini SE*) from the parallel run. The coefficients of *ini SE* for the unemployed and employed series are 0.21 and 0.18 respectively, and suggest the hybrid option still retains about 20% of the

errors in the final estimates after 12 months of changeover regardless of the quality of *ini SE*.

5.11 SE of final estimates regressing on ini SE

	Unemployed	Employed
Intercept	0.99578 (0.05123)	0.169087 (0.009134)
ini SE	0.20939 (0.01004)	0.180600 (0.011918)
Null deviance	0.8002000	0.0069333
Residual deviance	0.0090932	0.0001187
R^2	0.988636	0.98288

From table 5.10 and 5.11, we can see the consistent results for both unemployed and employed. Although the *ini SE* figures are not exactly equal to *Revision* plus *SE of final estimates*, we can confidently conclude, from the coefficients of *ini SE*, that the hybrid option reduces the errors by 80% over the 12 months after the changeover regardless the quality of the *ini SE*s. The remaining approximately 20% of errors is still likely in the final estimates.

6. DISCUSSION AND FURTHER WORK

This paper presents a set of SSM models and evaluations for a range of options for measuring statistical impacts from a survey redesign, using the ABS Labour Force Survey redesigns as a case study. The paper has shown that by modelling the differences of the GREG estimates for control and treatment groups at rotation group level, the model for measuring statistical impact from a parallel run simplifies the conventional SSM intervention analysis. This proposed model should be more robust and has the following advantages over the conventional SSM intervention approach:

- Eliminating the possible complications due to modelling the "true" population during parallel run periods;
- Avoiding the smoothing effect because of the lagged composite weights;
- Taking account of the dynamics of sample rotation induced process more effectively.
- Initialising the Kalman filter with a-priori information rather than diffuse prior because the model is stationary and the expected variances (covariances) of states are used to speed up the Kalman filter convergence rate.

Through theoretical deliberation and empirical simulation study, we now understand the relationship between the precision of detecting a statistical impact and:

1. the parallel run parameters, (i.e. the intra-cluster correlation between the treatment and control groups, the sample size, duration of parallel run);

2. the effect of sampling error, (i.e. the correlation structure of statistical impacts at rotation group level, and how it affects the statistical impact to the final composite estimate), and
3. the improvement and revision properties of the hybrid option after changeover.

The knowledge presented in this paper articulates how the survey parallel parameters, the characteristics of the LFS survey, such as intra-cluster correlation and sampling errors, and the properties of the SSM affect the precision of an estimated statistical impact. Therefore, it can assist in the decision making process for managing the risk, quality and cost of a statistical impact measurement strategy.

In terms of the options considered for measuring impact and implementing change, it is clear that a scenario without a parallel run is relatively inexpensive but has major draw backs in terms of the risk around the quality of published time series data (particularly coherence and interpretability) during the changeover period. In addition, the required accuracy criterion is unlikely to be achievable with this approach. For a critically important survey like the LFS, a large scale parallel run is required (assuming a low appetite for accepting statistical impacts on the series).

There are two possibilities to reduce the costs of the parallel run. Either the precision goal that an impact of one standard error must be detectable is relaxed or revisions of the estimated impact must be acceptable. In the latter case the time series modelling approach can be combined with a smaller sample size for the parallel run as illustrated with the six different scenarios investigated in Section 5.

For small parallel runs, there is of course, a large risk that the revision of the initial estimates for the SIM is substantial because the small parallel run does not produce precise initial estimates. This implies that the decision for making the changeover is based on an imprecise initial estimate. In a worst-case scenario the initial SIM estimates suggest a small impact and 12 month after the changeover the final SIM estimates appear to be substantially larger. This risk of course declines with the size of the parallel run and can be visualized by looking at the ratio of the revision and the standard error of the final SIM estimates as illustrated in Figure 5.4.

Our study of the hybrid option suggests that useful information obtained from the SIM in Phase 1 activities, such as small experiments, field tests and dress rehearsals, can be used as priors for the SSM of a parallel run. In other words, using SSM modelling approach, the SIM information obtained from a current phase can be used as the priors and input to the SSM of the next phase. The SIM precision can be continually improved over the three phases.

Further work is required to build on the learnings from this study as described in this paper. Some areas for future consideration may include:

- Use of other data sources: The SSM model (eqn. 2.1 – 2.5) used for the hybrid option can be extended to include related data sources in a multivariate seemingly unrelated time series equation (SUTSE) model to improve SIM precision by better predicting the true population. For example, Zhang and Honchar (2016) used unemployment benefit claimant counts (CC) is such related series for LFS unemployment, and the ANZ job advertisement (ANZadv) and Department of Employment Internet Job Vacancy index (DoEIVI).
- Evolving level shifts: In this paper we present results for a fixed level shift type of statistical impact. In reality the statistical impact induced by statistical program change could be evolving over time. Therefore, the current fixed intervention analysis needs to be extended to deal with an evolving level shift type of statistical impacts with, for example, a random walk model. We should at least use the evolving level shift model to test if the intervention is fixed.
- Kalman filter initialisation: The study presented in this paper demonstrated that the performance of the parallel run SSM can be further improved with a-priori information in an exact initialization of the Kalman. Therefore, it is recommended that further study is needed to maximise the precision of a parallel run / intra cluster correlation by taking advantage of statistical methods from the theory of randomized experiments. Such methods would have broad application for SIM phases 1 and 2. With a good understanding of its properties, particularly in comparison to the SSM based approach, we can position this alternative approach properly for SIM.
- Alternative SSM formulation: Further study is needed to explore alternative SSM model formulations which may utilise the historical data better to improve the Kalman filter convergence rate for shortening parallel run periods and to reduce the standard error of the estimated statistical impact.

ACKNOWLEDGEMENTS

The authors are particularly thankful to Dr. Siu-Ming Tam, Sybille McKeown, Prof. James Brown, Kristen Stone, Annette Kelly, Rosalynn Mathews, Bruce Fraser, Jacqui Jones and Bjorn Jarvis for their constructive and valuable suggestions and comments. 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) (2016) *Labour Force, Australia, April 2016*, cat. no. 6202.0, ABS, Canberra.
<http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/6202.0Explanatory%20Notes1Apr

%202016?OpenDocument>

Aoki, Masanao (1987). *State Space Modelling of Time Series*. Springer-Verlag Berlin Heidelberg.

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

Koopman, S.J.; Shephard, N.. and Doornik, J.A (2008) *SsfPack 3.0 Statistical Algorithems for Models in State Space From*. Timberlake Consultants Ltd., London.

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,* 16 339-348

Van den Brakel, J.A. (2008) "Design-based Analysis of Embedded Experiments with Applications in the Dutch Labour Force Survey", *Journal of the Royal Statistical Society, Series A*, 171, pp. 581–613.

Van den Brakel, J.A. 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.A. 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.A.; 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.

Van den Brakel, J.A., G. Griffiths, T. Surzhina, P. Wise, J. Blanchard, X. Zhang and O. Honchar (2017). A framework for measuring the impact of transitions in official statistics. *Australian Bureau of Statistics, Research paper* 1351.0.55.158, Canberra, Australia.

Van den Brakel, J.A. and R.H. Renssen, (2005). Analysis of experiments embedded in complex sampling designs, *Survey Methodology*, vol. 31, nr. 1, pp. 23-40.

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*

APPENDIXES

Appendix 1. Sample size needed to test one mean and the Minimum Detectable Impact (MDI) ratio

1.1 Calculate sample size needed to test one mean: one-sample, two-sided equality

The formulas below are useful for tests concerning whether a mean, u, is equal to a reference value, u_0. The Null and Alternative hypotheses are

$$H_0 : u = u_0$$
$$H_1 : u \neq u_0$$

Formulas

$$n = \left(\sigma \frac{z_{1-\alpha/2} + z_{1-\beta}}{u - u_0} \right)^2 \qquad \text{(A1.1)}$$

$$1 - \beta = \Phi(z - z_{1-\alpha/2}) + \Phi(-z - z_{1-\alpha/2}) \quad , \quad z = \frac{u - u_0}{\sigma / \sqrt{n}} \qquad \text{(A1.2)}$$

where
n is sample size
σ is standard deviation, $\sigma / \sqrt{n}$ is an estimator of the standard error of u
Φ is the standard Normal distribution function
Φ^{-1} is the standard Normal quantile function
α is Type I error
β is Type II error, meaning 1−β is power

1.2 Type I and Type II Error rates and the Minimum Detectable Impact (MDI) ratio

Consider the case where we have an estimate $\widehat{\Delta}$ of the statistical impact Δ arising from a change in measurement procedure and the standard error on $\widehat{\Delta}$, $SE(\widehat{\Delta})$ is estimated to be $\widehat{SE}(\widehat{\Delta})$.

We assume that the test statistic we will use $z = \frac{\widehat{\Delta}}{\widehat{SE}(\widehat{\Delta})}$ has a $N(0,1)$ distribution under H_0: $\Delta = 0$ and a $N(\frac{u}{SE(\widehat{\Delta})}, 1)$ distribution under H_1: $\Delta = u \neq 0$ and that we will declare there to have been a statistical impact (reject H_0) if $|z| > z_{1-\alpha/2} = \Phi^{-1}(1 - \alpha/2)$ where Φ is the standard Normal distribution function and α is the prescribed Type I error rate (probability of declaring an impact when there is none).

If β is the prescribed maximum Type II error rate when $\Delta \geq u$ (maximum probability of declaring no impact when the true impact is $\Delta \geq u$)

Then $\beta = \underset{d \geq u}{Max}(\text{Prob}(|z| \leq z_{(1-\alpha/2)} \mid \Delta = d)) = \Phi\left(z_{1-\alpha/2} - \frac{u}{SE(\hat{\Delta})}\right) - \Phi\left(-z_{1-\alpha/2} - \frac{u}{SE(\hat{\Delta})}\right)$ and given α and β we may solve this equation for $\frac{u}{SE(\hat{\Delta})}$. In the case where $\alpha = 0.05$ and $\beta = 0.5$ we obtain the solution $\frac{u}{SE(\hat{\Delta})} = 1.96$. Note that a corollary of the formula for β is that sample sizes meeting the 5% type I error will automatically satisfy a power requirement of 50%.

In terms of a standard measure of accuracy relevant to the standard error $SE(\hat{\text{L}})$ of an original population level estimate (e.g. employed or unemployed of the LFS) $\hat{\text{L}}$, we define the ratio of minimum detectable impact (MDI) as $\frac{u}{SE(\hat{\text{L}})}$ and see that to meet requirements we need:

$$\frac{u}{SE(\hat{\text{L}})} \geq 1.96 \frac{SE(\hat{\Delta})}{SE(\hat{\text{L}})}$$

Hence if we choose $u = SE(\hat{\text{L}})$ and require a 5% Type I error and Type II error of 50% (see the accuracy criterion in Section 3), then our measurement design will need to ensure that $1.96 \frac{SE(\hat{\Delta})}{SE(\hat{\text{L}})} \leq 1$ (or $\frac{SE(\hat{\Delta})}{SE(\hat{\text{L}})} \leq \frac{1}{1.96} = 0.51$).

APPENDIX 2. ESTIMATION OF INTRA CLUSTER CORRELATION OF EMPLOYED AND UNEMPLOYED

2.1 Estimates generation from the historical LFS data

The intra cluster correlation between treatment and control group sampling errors is a crucial parameter that influences the power of a parallel run to detect a statistical impact in a time series. However, this intra cluster correlation is not usually known a-priori. An exercise was conducted to approximately estimate its value from the historical LFS data by splitting the historical sample for each month into two subsamples then estimating the intra cluster correlation between the sampling errors of the two subsamples.

First, each cluster of dwellings in the last month of estimation period (in this study – November 2016) was systematically split into two half-clusters (A and B), i.e. using the current LFS selection method but with twice the skip than was actually used in the LFS. Then we matched backwards, that is, the most recent month (November 2016) had the entire responding sample allocated to subsamples A and B as above and this month was matched to the previous month (October 2016). We then continued to work backwards and each time we did a pairwise match of a given month with the previous month. This split used Primary Sample Unit (PSU) and time as the block variables in term of experimental design.

The GREG estimates for employed and unemployed persons were estimated from the split sample sets for all eight waves at the rotation group level by calibrating the split sample GREG estimates to the population characteristics. The period of time used for correlation calculation in this study was from January 2006 till November 2016.

2.2 A simple intra cluster correlation estimator

Let $\begin{pmatrix} y_{1,w,t} \\ y_{2,w,t} \end{pmatrix}$ be the two estimates from a split sample (control and treatment) of a rotation group for wave *w*. y_t is the current published LFS composite estimate for population y_t of total employed and unemployed persons in original terms. Since there is no significant rotation group (RG) effect in $\hat{y}_t$, we can assume $\begin{pmatrix} e_{1,w,t} \\ e_{2,w,t} \end{pmatrix} = \begin{pmatrix} y_{1,w,t} - \hat{y}_t - b_w \\ y_{2,w,t} - \hat{y}_t - b_w \end{pmatrix}$ contains only sampling and rotation group induced error process for the two split samples for wave *w*, where b_w is the wave bias. $\begin{pmatrix} y_{1,w,t} \\ y_{2,w,t} \end{pmatrix}$ are multiplied by 8 to align with published LFS composite estimates.

$\begin{pmatrix} e_{1,w,t} \\ e_{2,w,t} \end{pmatrix} \sim iid(0, \Sigma_w)$ where $\Sigma_w = \begin{pmatrix} \sigma_{1,w}^2 & \rho_w \sigma_{1,w} \sigma_{2,w} \\ \rho_w \sigma_{1,w} \sigma_{2,w} & \sigma_{2,w}^2 \end{pmatrix}$

It can be shown that $\rho = \rho_w, \quad \forall\ w$.

$$\hat{\rho}_w = \sum_{t=1}^{T} e_{1,w,t} \times e_{2,w,t} \Big/ \sqrt{\sum_{t=1}^{T} e_{1,w,t}^2 \sum_{t=1}^{T} e_{2,w,t}^2}$$

$$\hat{\rho} = \frac{1}{8} \sum_{w=1}^{8} \hat{\rho}_w$$

2.3 A SSM intra cluster correlation estimator

The purpose of this note is to derive a method / model to produce an estimate of cross-correlation between control and treatment groups sampling errors using the historical LFS data.

Assume $\hat{y}_{i,t} = \begin{pmatrix} \hat{y}_{1,i,t} & \hat{y}_{2,i,t} \end{pmatrix}'$ is a vector of GREG estimates for "control" and "treatment" samples (related to described above subsamples A and B). Then the model for $\hat{y}_{i,t}$ can be written as

$$\begin{pmatrix} \hat{y}_{1,1,t} \\ \hat{y}_{2,1,t} \\ \vdots \\ \hat{y}_{1,8,t} \\ \hat{y}_{2,8,t} \end{pmatrix} = 1_{[16]} y_t + \begin{pmatrix} b_1 \\ b_1 \\ \vdots \\ b_8 \\ b_8 \end{pmatrix} + \begin{pmatrix} e_{1,1,t} \\ e_{2,1,t} \\ \vdots \\ e_{1,8,t} \\ e_{2,8,t} \end{pmatrix} \quad \text{(A2.1)}$$

were the signal is y_t, the rotation group bias $b_1,...,b_8$ and the sampling error components can be expressed by equations (2.1-2.5) from section 2, with the only difference being that the sampling error components here have double dimensions reflecting the fact that two sub-samples for each wave have different sampling error component (although the coefficients of AR model and sampling error variance are the same for the two splits).

For any wave i, the sample error covariance matrix is

$$\Sigma_{e_i} = \begin{pmatrix} \sigma^2_{e_{1,i}} & \sigma_{e_{1,i}}\sigma_{e_{2,i}}\rho \\ \sigma_{e_{1,i}}\sigma_{e_{2,i}}\rho & \sigma^2_{e_{2,i}} \end{pmatrix}, \quad \text{(A2.2)}$$

where ρ is intra cluster correlation between treatment and control group sampling errors.

The relationship of sampling error variance and sampling error disturbance variance can be described by an equation

$$\sigma^2_{u_i} = \gamma\sigma^2_{e_i}, \quad \text{(A2.3)}$$

where the covariance of the corresponding sampling error disturbance is

$$\Sigma_{u_i} = \begin{pmatrix} \sigma^2_{u_{1,i}} & \sigma_{u_{1,i}}\sigma_{u_{2,i}}\rho \\ \sigma_{u_{1,i}}\sigma_{u_{2,i}}\rho & \sigma^2_{u_{2,i}} \end{pmatrix} \quad \text{(A2.4)}$$

and a loading factor is defined as

$$\gamma_i = \begin{cases} 1, & i=1 \ (\phi_1 = 0, \phi_2 = 0) \\ 1-\phi_1^2, & i=2 \ (\phi_1 \neq 0, \phi_2 = 0) \\ (1+\phi_2)[(1-\phi_2)^2 - \phi_1^2]/(1-\phi_2), & i \geq 3 \ (\phi_1 \neq 0, \phi_2 \neq 0) \end{cases} \quad \text{(A2.5)}$$

If the purpose is to use the rotation group level GREG estimate to estimate the correlation between the sampling errors of control and treatment samples, the simplest approach is to use the rotation group level GREG estimate from wave one because the additional GREG estimates from the other waves do not add extra value but increase the complicity of the model specification.

Therefore the simplified model for estimation correlation between the sampling errors of control and treatment samples includes only two variables in the measurement equation

$$\begin{pmatrix} \hat{y}_{1,1,t} \\ \hat{y}_{2,1,t} \end{pmatrix} = 1_{[2]} y_t + \begin{pmatrix} e_{1,1,t} \\ e_{2,1,t} \end{pmatrix} \tag{A2.6}$$

Note that the RGB component disappeared in the equation (A2.6) because only wave one is involved and therefore the RGB will be absorbed by signal y_t.

The cross correlation of the sampling error of the control and treatment samples can be estimated from the above model with the sampling error disturbance covariance matrix (A2.4) (loading factor $\gamma = 1$).

2.4 Results of correlation calculation

The following shows the results of the cross correlation estimates with and without the wave bias:

Wave bias parameter values:

Wave	1	2	3	4	5	6	7	8
Employed	4.657e-3	0.476e-3	0.763e-3	0.557e-3	1.078e-3	0.954e-3	0	0.240e-3
Unemployed	8.508e-2	4.917e-2	3.833e-2	2.722e-2	0.774e-2	1.539e-2	0	-1.113e-2

Intra cluster correlation with wave bias correction:

Wave	1	2	3	4	5	6	7	8	Overall
Employed	0.5759	0.3651	0.5447	0.4354	0.5513	0.5293	0.5412	0.5253	0.5085
Unemployed	0.1293	0.1648	0.2017	0.0965	0.0940	0.0936	0.1846	0.1856	0.1437

Intra cluster correlation without bias correction:

Wave	1	2	3	4	5	6	7	8	Overall
Employed	0.5759	0.3651	0.5447	0.4354	0.5513	0.5293	0.5412	0.5253	0.5085
Unemployed	0.1293	0.1646	0.2016	0.0964	0.0940	0.0936	0.1846	0.1856	0.1437

The inclusion of the wave bias made virtually no difference to the aggregate estimate of the cross correlation.

The estimated from SSM correlation is:

1) for employment: 55.6%,

2) for unemployment: ~ 0%.

APPENDIX 3: SIMULATED DATA GENERATION

Data generation description

The observations from the simulated data set followed the following structure,

$$\hat{y}^{g}_{i,t} = y_t + b^g_i + e^g_{i,t}$$

and sampling error for wave $i \geq 3$ and $t \geq 3$

$$e^g_{i,t} = \phi_1 e^g_{i-1,t-1} + \phi_2 e^g_{i-2,t-2} + \delta^g_{i,t} \qquad \delta^g_{i,t} \sim NID(0, \sigma^2_{i,\delta})$$

Note: The above equations are referenced from equation 2.1, 2.3 to 2.5

where,
$i \in (1, 2, ..., 8)$ is the wave index
$g \in (1, 2)$ is the group index where 1=control group and 2=treatment group

t is the time period

y_t is the "true" population estimate used in the simulation for employment and unemployment at time t

b^g_i is the rotation group bias (RGB) for the i^{th} wave of the control and treatment group, note that the rotation group bias is time-invariant

$e^g_{i,t}$ is the sampling error for the i^{th} wave of the control and treatment group at time t. It follows an autoregressive process of order 2, AR(2) with the disturbance term $\mu^g_{i,t}$,which is normally and independently distributed.

"True" population estimate y_t

It was estimated by using state space models on the LFS national level estimate. The final estimate was obtained by excluding the standard error component in the state space model.

The rotation group bias b^g_i

Each wave had a predefined rotation group bias value subject to a specific employment and unemployment simulation scenario.

AR2 sampling error $e^g_{i,t}$

The structure of the errors in the wave level time series is serially dependent across waves. For the purposes of this simulation study, $e^g_{i,t}$ and $\delta^g_{i,t}$ were required to satisfy a predefined variance covariance structure subject to a specific employment and unemployment simulation scenario.

$e_{i,t}^{g}$ could then be generated by calculating $\phi_1 e_{i-1,t-1}^{g} + \phi_2 e_{i-2,t-2}^{g} + \delta_{i,t}^{g}$ in the following recursive process to reflect the 8 months rotation cycle.

For $t = 1, 2, 3 \ldots\ldots$
Sampling error for wave $i = \mathrm{mod}(t,8) = 1$, control and treatment group,

$e_{i,t}^{g} = \delta_{i,t}^{g}$ No AR process

Sampling error for wave wave $i = \mathrm{mod}(t,8) = 2$, control and treatment group,

$e_{i,t}^{g} = \phi_1 e_{i-1,t-1}^{g} + \delta_{i,t}^{g}$ AR1 process

Sampling error for wave $i = \mathrm{mod}(t,8) = 3,4,5,6,7,0$ and $t \geq 3$, control and treatment group,

$e_{i,t}^{g} = \phi_1 e_{i-1,t-1}^{g} + \phi_2 e_{i-2,t-2}^{g} + \delta_{i,t}^{g}$ AR2 process

Table A3.1 presents the some key parameters for both control and treatment samples.

A3.1 Parameters for Simulation data generation

	Employed	Unemployed
Sample size per month	30000	
Sampling error AR1 for wave 2	0.835	0.589
Sampling error AR2 for wave 3 to 8	0.585, 0.3	0.466, 0.208
	RGB Control	
RSE at RGB	0.94%	6.60%
b_1	0.007	0.058267930
b_2	0.001	0.019303798
b_3	-0.0044	0.006714512
b_4	-0.0044	0.000405143
b_5	0.0005	0.017966054
b_6	0.0001	0.019514134
b_7	0	0
b_8	0.0002	0.046400917
	RGB Treatment	
$b_1^{(n)}$	0.6	0.7
$b_2^{(n)}$	0.001	0.2
$b_3^{(n)}$	-0.001	0.2
$b_4^{(n)}$	-0.001	-0.02
$b_5^{(n)}$	0.001	-0.02
$b_6^{(n)}$	0	-0.02
$b_7^{(n)}$	0	0
$b_8^{(n)}$	-0.6	-0.68

The following pseudo code illustrates the data simulation process

```
Set RSE RG control sample

Iterate replicates 1 to 100
        Iterate parallel run duration:  11, 13, 15 19
                Iterate Kappa: 0.3. 0.5. 0.8, 1
                Derived Treatment sample RSE from Kappa.
                        Iterate intra cluster correlation: 0, 0.3, 0.5, 0.8
```

```
                Generate both control and treatment sample
            End
        End
    End
End
```

Note:

The simulation program required the input parameters, ϕ_1 for wave 2, ϕ_1 and ϕ_2 for wave 3 to wave 8, standard error of control and treatment group (reflects κ), ρ, and RGB control and treatment parameters.

In the data simulation, the AR1 parameter ϕ_1 for wave 2 was different to the AR1 parameter ϕ_1 for wave 3 and beyond.

A different seed was used to generate the white noise component in each replication and therefore, the simulated observations differed only in the sampling error component.

APPENDIX 4: THE STATE CORRELATION MATRIX AND THE STATISTIAL IMPACT TO THE OVERALL COMPSITE ESTIMATES

The SSM model equation (3.9) – (3.10) can rewritten in a condensed form as

Observation equation:

$$\mathbf{d}_t = \mathbf{F}\boldsymbol{\alpha}_t \qquad \text{(A4.1)}$$

State equation

$$\mathbf{X}_t = \mathbf{G}\mathbf{X}_{t-1} + \boldsymbol{\omega}_t \qquad \text{(A4.2)}$$

where $\mathbf{d}_t = \hat{\mathbf{y}}_t^{(\boldsymbol{\tau})} - \hat{\mathbf{y}}_t$, $\mathbf{X}_t = \begin{pmatrix} \boldsymbol{\alpha}_t \\ \mathbf{e}_t \\ \mathbf{e}_{t-1} \end{pmatrix}$, $\mathbf{F} = \begin{pmatrix} \mathbf{I} & \mathbf{I} & \mathbf{0} \end{pmatrix}$, $\boldsymbol{\omega}_t = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\delta}_\mathbf{t} \\ \mathbf{0} \end{pmatrix}$ $\quad \boldsymbol{\omega}_t \sim N(0, \Omega)$

The AR1 model (A4.2) of state vector $\mathbf{X}_t$ can be written as a MA form below

$\mathbf{X}_t = \mathbf{G}\mathbf{X}_{t-1} + \boldsymbol{\omega}_t = \sum_{i=0}^{\infty} \mathbf{G}^i \boldsymbol{\omega}_{t-i}$ and covariance matrix can be derived as

$$\boldsymbol{\Gamma}(0) = \left\langle \mathbf{X}_t \quad \mathbf{X}_t' \right\rangle = \sum_{i=0}^{\infty} \mathbf{G}^i \boldsymbol{\omega}_{t-i} \left(\sum_{i=0}^{\infty} \mathbf{G}^j \boldsymbol{\omega}_{t-i} \right)' = \sum_{i=0}^{\infty} \boldsymbol{\Theta}^{\mathbf{i}} \Omega \boldsymbol{\Theta}'^{\mathbf{i}} \qquad \text{(A4.3)}$$

Substituting the sampling error AR2 coefficients, and the diagonal covariance of sample error disturbance for Ω, we have the analytical solution for the correlation matrix $\boldsymbol{\rho}_{\alpha}$ correspond to state $\boldsymbol{\alpha}_t$. Table A4.1 and A4.2 shows the correlation matrix of state $\boldsymbol{\alpha}_t$

A4.1 State $\boldsymbol{\alpha}_t$ correlation matrix $\boldsymbol{\rho}_{\alpha}$ of LFS unemployed

Wave	1	2	3	4	5	6	7	8
1	1.000	0.589	0.483	0.348	0.262	0.195	0.145	0.108
2	0.589	1.000	0.589	0.483	0.348	0.262	0.195	0.145
3	0.483	0.589	1.000	0.589	0.483	0.348	0.262	0.195
4	0.348	0.483	0.589	1.000	0.589	0.483	0.348	0.262
5	0.262	0.348	0.483	0.589	1.000	0.589	0.483	0.348
6	0.195	0.262	0.348	0.483	0.589	1.000	0.589	0.483
7	0.145	0.195	0.262	0.348	0.483	0.589	1.000	0.589
8	0.108	0.145	0.195	0.262	0.348	0.483	0.589	1.000

Lagend

lag 0	lag ± 1	lag ± 2	lag ± 3	lag ± 4	lag ± 5	lag ± 6	lag ± 7

A4.2 State $\boldsymbol{\alpha}_t$ correlation matrix $\boldsymbol{\rho}_{\alpha}$ of LFS employed

Wave	1	2	3	4	5	6	7	8
1	1.000	0.835	0.788	0.712	0.653	0.595	0.544	0.496
2	0.835	1.000	0.835	0.788	0.712	0.653	0.595	0.544
3	0.788	0.835	1.000	0.835	0.788	0.712	0.653	0.595
4	0.712	0.788	0.835	1.000	0.835	0.788	0.712	0.653
5	0.653	0.712	0.788	0.835	1.000	0.835	0.788	0.712
6	0.595	0.653	0.712	0.788	0.835	1.000	0.835	0.788
7	0.544	0.595	0.653	0.712	0.788	0.835	1.000	0.835
8	0.496	0.544	0.595	0.653	0.712	0.788	0.835	1.000

Lagend

lag 0	lag ± 1	lag ± 2	lag ± 3	lag ± 4	lag ± 5	lag ± 6	lag ± 7

For the LFS unemployed, it can be seen that the state correlation matrix is induced by the serial cross-correlation of the wave sampling errors (AR(1) parameter - AR11 = 0.589 for wave 2; AR(2) parameters AR12 = 0.466 AR22=0.208 for wave 3-8. The coloured legends represent the lagged cross-correlation of the sampling error

In other words, the state correlation matrix is the same as the sum of serial cross-correlations of different waves (or the sum the cross-correlation matrix of all orders/lags) of the sampling errors induced by the AR(2) process[17].

[17] Similarly, we also can show that the element of sampling error state are concurrently independent, ie. its

The special analytical results are crucially important in two aspects that
(1) they can be used as a part of state prior information for Kalman filter initialisation and
(2) they also demonstrate that the estimated wave level statistical impact $\{\alpha_i\}$ are concurrently correlated while the sampling error state $\{e_i\}$ are concurrently independent.

The variance of the overall impact to LFS composite estimates can be estimated by $\frac{1}{8^2}\mathbf{W}^T\mathbf{P}_t\mathbf{W}$ where $\mathbf{P}_t$ is the covariance matrix of state $\boldsymbol{\alpha}_t$, $\mathbf{W}$ is the concurrent composite weight vector. Let $\boldsymbol{\sigma}_{\boldsymbol{\alpha}}$ denoted the standard error vector of $\boldsymbol{\alpha}_t$, then $\mathbf{P}_t = \boldsymbol{\sigma}_{\boldsymbol{\alpha}}\boldsymbol{\rho}_{\boldsymbol{\alpha}}\boldsymbol{\sigma}_{\boldsymbol{\alpha}}'$. When the elements of $\mathbf{W}$ are close to 1, and all the elements of $\boldsymbol{\sigma}_{\boldsymbol{\alpha}}$ are the same or very similar, then $\frac{1}{8^2}\mathbf{W}^T\mathbf{P}_t\mathbf{W} \simeq \frac{1}{8^2}\left(\mathbf{1}_{[8]}\right)'\boldsymbol{\rho}_{\boldsymbol{\alpha}}\mathbf{1}_{[8]}\sigma_\alpha^2$. Therefore, standard error of the overall impact to composite estimate can be derived from the wave level standard error with a multiplier $\sqrt{\frac{1}{8^2}\left(\mathbf{1}_{[8]}\right)'\boldsymbol{\rho}_{\boldsymbol{\alpha}}\mathbf{1}_{[8]}}$. 0.6798 and 0.8681 are the multiplier values for LFS unemployed and employed respetively. If the elements of the state $\boldsymbol{\alpha}_t$ are inapproperiately assumed independent, ie. $\boldsymbol{\rho}_{\boldsymbol{\alpha}} = \mathbf{I}$, the multiplier value is 0.3536. The standard error of the overall impact to the composite estimates of LFS unemployed and employed are likely to be under estimated by factors of 0.520 and 0.407 respectively.

correlation matrix is an identical matrix, while they are temporally correlated.

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