Using Administrative Data for Seasonal Adjustment of Survey Time Series in the Presence of a Major Measurement Change

Xichuan (Mark) Zhang[1] and Nick Von Sanden
Australian Bureau of Statistics

Abstract: The introduction of a major change in the design of a repeated survey will often lead to a significant structural change in the time series output of a survey. The quality of seasonally adjusted estimates are in question when structural changes in the time series occur as a result of methodological changes (e.g. changes in sample design or estimation) which are not related to any changes in the real economy.

Typically, to help assess the potential structural changes in the time series, the survey is subject to a "parallel run" where the old methodology and the new methodology are applied simultaneously for a number of overlapping periods. By analysing the outputs from the parallel run, the analyst can estimate the effects of the change on the level and/or seasonality of the survey data and make appropriate adjustments by revising the historical time series before seasonal adjustment. These parallel run schemes are, however, very expensive to conduct and there is a strong desire to reduce the scale of this exercise, e.g. by restricting the number of periods that are subject to the parallel run.

In this research, we investigated using time series data sourced from administrative records to estimate the change to seasonal patterns resulting from a methodological change. We illustrate our method using data from an upcoming redesign of the Australian Quarterly Business Indicator Survey; which is being redesigned to introduce a new industry classification. By modelling the survey data and time series information from Business Activity Statements, we show it is possible to measure the structural change in seasonality involved, and reduce the number of periods of parallel runs to a minimum that is needed for establishing the change in level of the time series.

Key words: Time Series, Structural Change, Seasonality, Administrative Record

1. Introduction

Official time series statistics are widely applied to analyse the current state and direction of an economy. Major methodological changes in surveys, used to produce time series statistics, will reduce the comparability of the data over time as estimation methodologies and definitions may change. Inconsistent time series are a major problem in economic research because accurate analysis and forecasting relies on the inter-temporal comparability of the time series estimates.

Overlapping parallel runs on the old and new basis (i.e. before and after the methodological change) provide a direct instrument to measure the impact of the methodological change. Long overlapping parallel run periods are usually very resource intensive. Official statistical agencies may not have sufficient resources to conduct parallel run periods of sufficient length.

In this paper we consider how available administrative data may be applied to estimate the impact of methodological changes on the seasonal pattern of time series estimates in a more timely and cost effective fashion. In section 2, we review some common methods in the literature. In section 3, we propose a general method to make an impact assessment of seasonal pattern changes by utilising related administrative data. An application is then presented in section 4 on the Australian Bureau of Statistic's (ABS) Quarterly Business Indicators Survey (QBIS) for which a time series of related

[1] This paper does not necessarily represent the considered views of the Australian Bureau of Statistics. Readers interested in the subsequent development of the research topic are encouraged to contact the author.

administrative data is available. A major change to the industry classification, and the associated survey estimation methodologies will be implemented to this survey in 2009. Section 5 is our concluding remarks.

2. Literature review:

Issues associated with structural changes to time series, induced by methodological changes to surveys, have been of concern to official statistical agencies for many years. For example, Bureau of Economic Analysis (1993) explains the impacts of methodological changes on time series and outlines why the impact cannot be overlooked during critical planning stages. It is recognised that there are competing views among statistical agencies on the importance of time series continuity and consistency depending upon the type of data these agencies produce.

There is no standard approach in dealing with structural changes in time series. Different approaches are applied depending on the cost and the level of detail available regarding the old and new survey methodologies. A "standard" approach to preserving continuity after introducing a new survey methodology is to create a linkage where the series structural change occurs. This is accomplished by producing the time series on both the old and new survey methodology for an overlapping transition period. With the dual estimated results, a full impact of the new survey methodology can be assessed at aggregate levels. The measured impact then can be used to backcast the old series and produce a revised historical series which is consistent with the new series.

For sub-annual surveys, a major challenge is the unknown seasonal behaviour of the new time series. A convenient assumption is often made that the seasonal patterns of the old and new time series are the same. Under this assumption only a short period of overlapping parallel runs would be required to estimate a level shift impact.

In practice this assumption is not necessarily realistic for some methodological changes. When this assumption is violated and seasonally adjusted estimates must be released concomitantly with the first publication of the new time series, biases or distortions will be introduced into seasonally adjusted estimates because seasonal adjustment procedures, such as X-12-ARIMA, TRAMO/SEAT, require a consistent time series as input.

Time series model:

Most economic time series follow the multiplicative decomposition model (1):

$$O_j(t,l)=T_j(t,l)\times S_j(t,l)\times I_j(t,l) \quad , \quad \sum_{l=1}^{P} S_j(t,l)=P \tag{1}$$

where $j=1,2$ denotes the old and new survey methodology, $T_j(t,l), S_j(t,l)$ and $I_j(t,l)$ are trend-cycle, seasonal, and irregular components of the original estimate $O_j(t,l)$ on period *l* of year *t* respectively, $l=1,2,\dots,P$. . *P* is the periodicity of the time series. e.g. *P* = 4 and 12 for quarterly and monthly time series respectively.

Direct approach:

As mentioned earlier, overlapping parallel runs on the old and new survey methodologies provide original estimates $O_j(t,l)$ ($j=1,2$) while the estimates of $T_1(t,l)$ and $S_1(t,l)$ are available. The new survey methodology impact in terms of a level shift and seasonal outliers can be estimated by model (2):

$$f: \begin{cases} T_2(t,l)=\alpha\times T_1(t,l) \\ S_2(t,l)=\beta(l)\times S_1(t,l), \quad \sum_{l=1}^{P}\beta(l)=P \end{cases} \quad (2)$$

where α and $\beta(l)$ $(l=1,2,\ldots,P)$ are constant factors represent the interruptive level shift and seasonal pattern change. This model is often referred to as a Seasonal Level Shift (SLS) outlier model in the literature. It is a combination of a seasonal outlier along with a level shift.

Traditional time series intervention analysis for seasonal structural changes, see Kaiser, R. and Maravall, A. (2001), requires more than 3 years of observations of the new time series to accurately identify the interruptive distortion introduced by the new survey methodology. It may not be acceptable for an official statistical agency to conduct parallel runs and not publish seasonally adjusted estimates of the new series for three years. To reduce the length of parallel run required and solve the problem of the short span of this time series, McLaren and Zhang (2003) developed a method using a mixed model framework to estimate seasonal structural changes for an aggregate time series based on its disaggregated series. This method stacks the disaggregated series and estimates their common seasonal pattern change for the new time series against the old. The estimated common seasonal pattern change provides statistical inference to determine if a seasonal structural change is introduced by the new survey methodology. This method needs only one year and one further period of overlapping parallel estimates based on the old and new survey methodologies and consequently a substantial reduction in the required length of parallel run is achieved. Gatto (2006) developed a "heuristic" method using a regression framework to estimate the seasonal impact of a new survey methodology and also used a minimum of one year and one extra period of overlapping parallel estimates.

There is no way to estimate seasonal structural changes with less than one year and one extra period of parallel run without utilising further related information. In practice it may still be too expensive to conduct the minimum one year and one extra period of overlapping parallel runs and in this case our only option is to try and estimate the seasonal structural change indirectly with other related information.

3. Indirect impact assessments and their time series model

A cost effective estimate of the impact of methodological changes on a survey can be prepared indirectly by considering related administrative times series data. We now consider the properties of the related data in order to aid in seasonal pattern estimation for the survey data on the new basis.

Let us extend the time series model (1) including related information

$$O_{i,j}(t,l)=T_{i,j}(t,l)\times S_{i,j}(t,l)\times I_{i,j}(t,l), \quad \sum_{l=1}^{P}S_{i,j}(t,l)=P \quad (3)$$

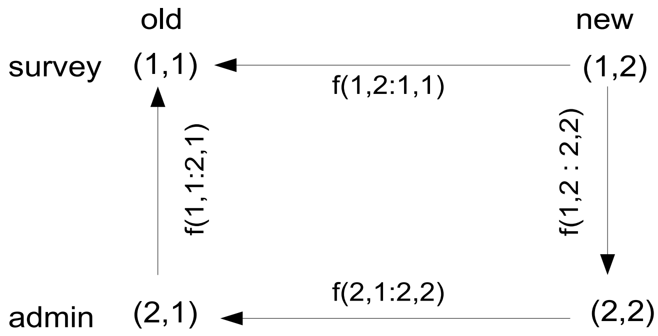
where $i=1,2$ denotes the survey data and the related administrative data and $j=1,2$ denotes the old and new survey methodologies respectively. Model equations (2) can be extended to

$$\boldsymbol{f(i,j:m,n)}: \begin{cases} T_{i,j}(t,l)=\alpha_{(i,j:m,n)}\times T_{m,n}(t,l) \\ S_{i,j}(t,l)=\beta_{(i,j:m,n)}(l)\times S_{m,n}(t,l), \quad \sum_{l=1}^{P}\beta_{(i,j:m,n)}(l)=P \end{cases} \quad (4)$$

where $i,j,m,n=1,2$, $i\neq m$ or $j\neq n$. Subscript $(i,j:m,n)$ denotes the relationship from (m,n) to (i,j).

The following instrument chart can describe the direct and indirect approaches.

Figure 1: Instrument chart



The top part illustrates the "standard" direct impact measure, $f(1,2:1,1)$, of comparing the new methodology to the old methodology over the overlapping period of a parallel run. i.e. Equation (1) can be written as following equation (5).

$$\begin{cases} T_{1,2}(t,l)=\alpha_{(1,2:1,1)}\times T_{1,1}(t,l) \\ S_{1,2}(t,l)=\beta_{(1,2:1,1)}(l)\times S_{1,1}(t,l) \end{cases} \quad (5)$$

Suppose the administrative data contains equivalent information to the survey under both the old and new methodologies. We can then derive an indirect approach by applying equation (4) in a clockwise fashion moving from new survey to old survey basis in figure 1 and obtain equation (6).

$$\begin{cases} T_{1,2}(t,l)=\alpha_{(1,2:2,2)}\times\alpha_{(2,2:2,1)}\times\alpha_{(2,1:1,1)}\times T_{1,1}(t,l) \\ S_{1,2}(t,l)=\beta_{(1,2:2,2)}(l)\times\beta_{(2,2:2,1)}(l)\times\beta_{(2,1:1,1)}(l)\times S_{1,1}(t,l) \end{cases} \quad (6)$$

The relationship $f(1,2:2,2)$ cannot be established before the new survey methodology is conducted for many periods. Assuming either $f(1,2:2,2)=f^{-1}(1,1:2,1)$ or $f(1,2:1,1)=f(2,1:2,2)$, we obtain

$$\begin{cases} T_{1,2}(t,l)=\alpha_{(2,2:2,1)}\times T_{1,1}(t,l) \\ S_{1,2}(t,l)=\beta_{(2,2:2,1)}(l)\times S_{1,1}(t,l) \end{cases} \quad (7)$$

from equation (6).

Although the administrative data may measure a similar concept to our survey estimates, relationship (7) may not necessary hold if the two time series are not equivalent. The assumption $\beta_{(1,2:1,1)}(l)=\beta_{(2,2:2,1)}(l)$ will still be justified in this case because the seasonal patterns are described as "standardised" factors which are not sensitive to the level of the series in the time series decomposition model (3). In comparison the indirect approach will not necessarily be appropriate for estimating the level shift. This is because the level shift from the old basis to the new basis survey could be different from the level shift in the administrative data. The next section provides some examples showing that the ratio between new and old time series derived from administrative records may not be a constant. In other words, we expect the seasonal pattern difference between survey estimates based on the new and old methodologies will be approximately the same as the seasonal pattern difference of the new and old administrative records while the level of difference of the new and old survey methodology may not necessary be close.

We therefore propose the following two-step method to reduce the length of parallel run required to estimate any seasonal pattern changes after a methodological change. Based on a reasonable seasonal pattern change estimate, we will be able to revise the historical time series and produce consistent seasonally adjusted estimates.

STEP 1: conceptually similar time series data is sourced from an administrative data set and the seasonal pattern change from the old to the new methodology is estimated based on this administrative time series.

STEP 2: a short parallel run is conducted (e.g. 2~3 periods) on the survey time series and the seasonal pattern change estimated from STEP 1 is applied. This allows us to produce a more appropriate estimate of the level shift from the old to the new method on the survey data.

The estimated seasonal pattern change (seasonal outlier) factors and level shift factors are then both applied to revise the historical time series and provide a consistent time series for seasonal adjustment. In the following section, we present an application of STEP 1 based on a real case study.

4. The Application

The ABS is currently preparing to implement the Australia New Zealand System of Industry Classification version 2006 (ANZSIC 06) to replace the existing 1993 version (ANZSIC 93) for all ABS surveys and the National Accounts. The initial strategy was to conduct 5-quarters (or 13-months) parallel runs to produce estimates for all ABS business sub-annual surveys under both ANZSIC 93 and ANZSIC 06. The sample size would be increased by 15% during this overlapping period to target those units which are not covered by ANZSIC 93. The impacts of ANZSIC 06 can be estimated for both level shift and seasonal pattern changes from these dual estimation results. The estimated impacts will then be used to revise (backcast) the historical time series and thereby maintain time series continuity for seasonal adjustment of ABS publication aggregates.

The ABS Quarterly Business Indicator Survey (QBIS) provides the following vital information on Australian businesses for twenty two ANZSIC 93 divisions and subdivisions since 2001.

(1) sales of goods and services,

(2) wages and salaries,

(3) profits, and

(4) inventories.

More details regarding this survey can be found in ABS (2001).

Due to competing priorities and the high cost involved, most ABS sub-annual business surveys are unlikely to carry out the initial strategy by conducting an extended parallel run period. It was decided that QBIS would only perform a parallel run for 2 overlapping quarters and a study was undertaken to see if related administrative data could be used to estimate ANZSIC 06 impacts on QBIS.

The Australian Taxation Office (ATO) collects Business Activity Statement (BAS) data for taxation administration purposes. This administrative record data set has been coded in both ANZSIC 93 and ANZSIC 06 since 2002 and 2005 respectively. The BAS data contains business turnover, wages, capital expenditure and non-capital expenditure. BAS Unit Record Estimates (BURE) are reporting estimates from the raw BAS data. Quarterly BURE is available from September quarter 2002. The BURE is not a timely indicator because the first vintage estimate for a specific reference period will only be available after 2 to 3 quarters, and its final vintage will not be available for at least for 5 quarters. i.e. the final BURE data is only produced a minimum of 5 quarters after the reference period. Due to this, BURE data has little value as a timely economic indicator.

The BURE ANZSIC 93 and BURE ANZSIC 06[2] data, used in this study, are available from September quarter 2002 to December quarter 2007 and from December quarter 2005 to December quarter 2007 respectively.

The estimates of QBIS and BURE are essentially separate measures of similar concepts, business labour costs (in term of wages and salaries) and turnover (in term of sales). Therefore, the BURE can be used as an indirect instrument to estimate the impact of seasonal pattern changes under ANZSIC 06.

Both ANZSIC 93 and ANZSIC 06 have four classification levels: Division, Subdivision (2 digits), Group (3 digits) and Class (4 digits). To demonstrate how STEP 1 of our proposed method works, we present a case study here to assess seasonal pattern changes in QBIS total Wages and its 14 industry division components.

We have developed a robust method[3] to estimate seasonal pattern differences between the old and new time series with minimum length of 5 quarters of overlapping data. This method does not rely on the assumption that the level shift between the old and new time series is a constant. This method has been applied to assess the impact on the seasonal pattern of moving from ANZSIC 93 to ANZSIC 06 on the BURE data. Results can be seen in Table 1.

Table 1: Total wages seasonal impact of ANZSIC 06 to ANZSIC 93

Period	$\beta_{(2,2:2,1)}(1)$	$\beta_{(2,2:2,1)}(2)$	$\beta_{(2,2:2,1)}(3)$	$\beta_{(2,2:2,1)}(4)$	**F-test p-value**
value	1.002	0.998	1.000	0.999	0.46
T-test p-value	0.23	0.33	0.56	0.57	

The value row shows the estimated seasonal impact as seasonal outlier factors for each quarter. The p-value of *t*-test and global p-value of *F*-test suggest that the seasonal impact is not statistically significant. In other words, we cannot reject the hypothesis that the seasonal patterns of BURE data under ANZSIC 93 and ANZSIC 06 are essentially the same.

An additional common seasonal pattern change test was also conducted based on the method of McLaren and Zhang (2003). Table 2 shows that the common seasonal pattern change of the BURE Wage total over its 14 components is not statistically significant either.

2 ANZSIC 06 BURE data actually start from June quarter 2005. It was found to have poor quality for the first two quarters. Therefore, June and September quarter 2005 ANZSIC 06 BURE are not used.

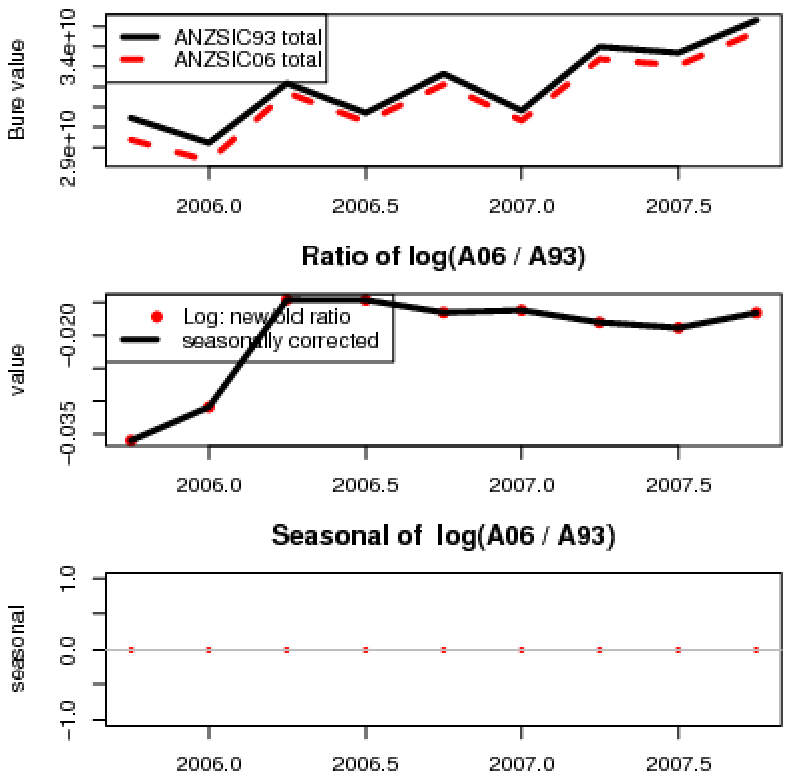
3 The details of the method is out scope of this paper. Interested readers can contact us for the technical details at mark.zhang@abs.gov.au.

Table 2: Common seasonal impact of ANZSIC 06 total wages to ANSIC 93

Period	**Quarter 1**	**Quarter 2**	**Quarter 3**	**Quarter 4**
value	1.000	1.005	0.998	0.997
***t*-test p-value**	0.98	0.66	0.87	0.71

Figure 2 provides graphical presentations for the BURE Wage total on ANZSIC 93 and ANZSIC 06 bases. A level shift between the two time series is evident in the top graph. The middle graph presents that the ratio between BURE ANZSIC 06 and BURE ANZSIC 93 on a logarithmic scale. It can be observed that the ratio is not necessarily constant over the time span examined. This is contrary to the assumption made in model (4). As mentioned earlier, a more robust procedure, which does not rely on the assumption of a constant level shift, is applied to extract an estimate of the impact on the seasonal pattern from the ratio. At a 10% statistical significance level we cannot reject the hypothesis that there are no ANZSIC 06 impacts on the seasonal pattern. The bottom graph shows that the seasonal outliers of the ratio between BURE ANZSIC 06 and BURE ANZSIC 93 are zeros on logarithmic scale. i.e. there is no statistically significant seasonal impact.

Figure 2: ANZSIC 06 seasonal pattern impact to BURE total wages



This result is not a surprise because we did not expect to see seasonal pattern changes for total wages as a whole. However, it does not necessarily mean that ANZSIC 06 will not have a seasonal pattern impact on the lower level components. Table 3 lists all fourteen industry divisions[4], *F*-test p-values for the possible ANZSIC 06 induced seasonal pattern changes at the division level.

Table 3: Tests of ANZSIC 06 seasonal pattern changes at division level

Description (in ANZSIC 93)	**ANZSIC 93 Industry division**	**ANZSIC 06 Industry division**	**F-test p-value**
Mining	B	B	0.076
Manufacturing	C	C	0.269
Electricity, Gas and Water Supply	D	D	0.231
Construction	E	E	0.100
Wholesale Trade	F	F	0.198
Retail Trade	G	G	0.472
Accommodation, Cafes and Restaurants	H	H	0.509
Transport and Storage	I	I	0.451
Communication Services	J	J	0.193
Finance and Insurance	K	K	0.154
Education	N	P	3.5e-05
Health and Community Services	O	Q	0.077
Cultural and Recreational Services	P	R	0.645
Personal and Other Services	Q	S	0.643
Total			0.464

At the 10% level there are significant seasonal pattern changes in four out of fourteen components. These divisions are Mining, Construction, Education and Health. In the following we use the Eduction division as an example to illustrate this seasonal pattern change.

The top graph of figure 3 shows time series of BURE wages for the education division on ANZSIC 93 and ANZSIC 06 bases. The level of BURE ANZSIC 06 series is always higher than the ANZSIC 93 series. This indicates that the scope of ANZSIC 06 is wider than ANZSIC 93. The ratio graph in the middle shows that the level shift is not a constant. The seasonally corrected ratio curve is much smoother than the original ratio. The estimated seasonal pattern changes are plotted in the bottom graph.

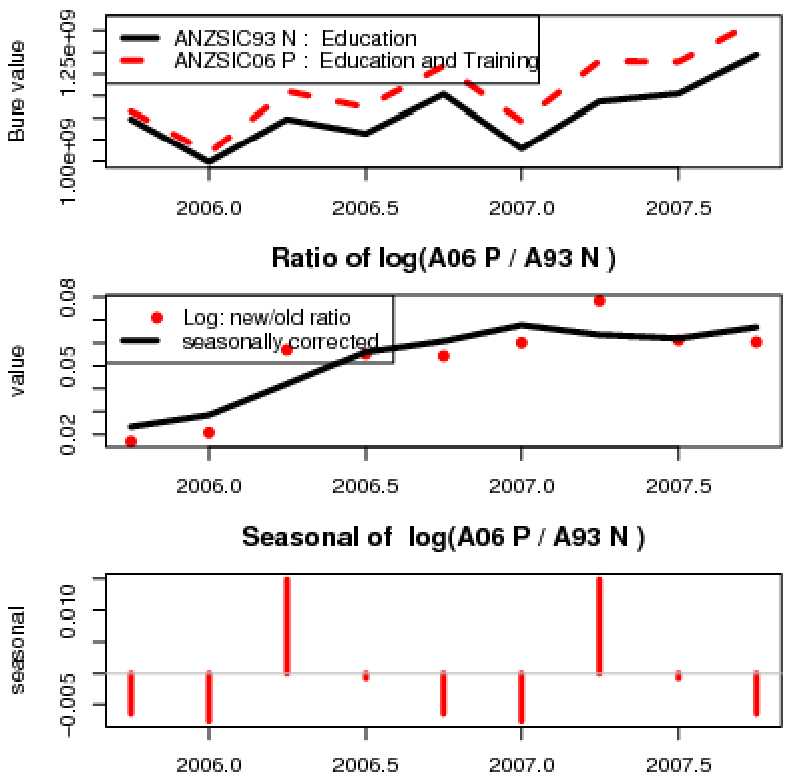
It can be seen clearly that the ANZSIC 06 seasonal pattern change is mainly in the second quarter (+1.5%) which is also compensated by the first and fourth quarters (-0.6% and -0.8%). This seasonal pattern change should be taken into account for the purposes of historical revision (backcasting) or seasonal adjustment after linking the ANZSIC 93 to ANZSIC 06 series. If this seasonal pattern change is not accounted for, the seasonally adjusted estimates would be distorted by this seasonal outlier and in this case a large proportion of the seasonal outlier would be likely to remain in the seasonally adjusted estimates. The possible consequences of not accounting for this seasonal pattern change could be

4 ANZSIC 93 business service industry division is excluded because it is split into three ANZSIC 06 divisions.

(1) a maximum of 1.5% and 2% biases for the seasonally adjusted estimates, on level and percentage movement respectively, at the current end of the linked series, and

(2) long lasting revisions to seasonally adjusted estimates of the historical part of the series due to the continual distortion to seasonal adjustment procedures relating to the presence of these seasonal outliers.

Figure 3: ANZSIC 06 Seasonal pattern impact to education division of BURE wages



A similar approach to the one presented above can be used on lower classification levels, i.e. for the division and its subdivision components, subdivision and group components, group and class components, as long as there exists a reasonable conceptual 1-to-1 mapping between ANZSIC 93 and ANZSIC 06 time series at the level. A potential problem is that these conceptual 1-to-1 mappings do not always exist for the lower level classifications. Therefore, the proposed method can only be used on certain higher classification levels which are usually the levels at which the ABS publishes.

Another challenge is that a survey item may not have an equivalent in the administrative data source. For example, QBIS surveys business profits while the BURE data does not have an item for profit. A potential solution would be to derive profits approximately from turnover less wages and expenditure. However, there is no reasonable method for deriving inventory estimates from the BURE data.

5. Concluding remarks

When a major measurement change occurs to a sub-annual survey, time series continuity (or inter-temporal comparability) between the old and new time series can be called into question. Structural changes to the time series create a challenging problem for official statistical agencies, such as the Australian Bureau of Statistics, who produce and publish seasonally adjusted and trend estimates concurrently. If structural changes are not assessed properly and accounted for appropriately because of the associated high cost, the timeliness and accuracy of key economic indicators can be cast in doubt.

This paper presents a practical solution to assessing measurement change by using administrative data sources to indirectly test and estimate structural changes to the seasonal pattern induced by a survey measurement change. This solution should be more practically feasible than using long periods of overlapping parallel runs, under both the old and new methodologies, in order to make a direct impact assessment of any seasonal pattern change. This proposed method can reduce the high costs associated with implementing a new methodology by reducing the length of parallel runs. In this way we can revise the historical time series to maintain time series continuity and produce quality seasonally adjusted and trend estimates in a timely and accurate fashion.

We recognise that the proposed method works under a very plausible assumption. However, whether the assumption accurately reflects the reality cannot be verified until a sufficient length (usually 3 years) of time series measured on the new basis becomes available. We believe that the proposed method is a feasible solution under the scenario that long periods of overlapping parallel runs are not affordable. Risks in using this proposed method need to be properly managed by good communication with data users and reassessing and refining the revisions to historical time series if necessary when sufficient data are available on the new basis.

Reference:

ABS (2001) Information Paper: Improvements to Australian Bureau of Statistics Quarterly Business Indicators, ABS Catalogue no. 5677.0. http://www.abs.gov.au.

ABS (2007) Information Paper: Quality Dimensions of the Australian National Accounts, ABS Catalogue no. 5216.0.55.002. http://www.abs.gov.au.

Bureau of Economic Analysis (1993) The Impact of Classification Revisions on Time Series, Issues Paper No. 5, Economic Classification Policy Committee, Bureau of Economic Analysis, U.S. Department of Commerce, Washington D.C. July 1993,

Gatto, Riccardo (2006), Series Revision and Seasonal Adjustment of Short Time Series in Presence of a Major Methodological Break, Conference On Seasonality, Seasonal Adjustment And Their Implications For Short-Term Analysis And Forecasting, Luxembourg

Kaiser, R. and Maravall, A. (2001) Seasonal outliers in time series, http://www.bde.es

McLaren, C. and Zhang, X. (2003), Estimation of seasonal factors for a short time span using multi-level modelling with mixed effects, the 54th Session of the International Statistical Institute, Berlin. Australian Bureau of Statistics Working Paper in Econometrics and Applied Statistics: No 2003/1 at http://www.abs.gov.au.