

RESEARCH REPORT ON THE BER BUSINESS TENDENCY SURVEYS

Laurie Binge

This draft: February 16, 2018

1 Introduction

This report presents the results from our research on the microdata from BER business tendency surveys. After cleaning the microdata, we calculate the balance statistics for all of the sectors and check that the microdata provides results that are consistent with the BER's published results. Next, we calculate new series for all of the questions, for the main sectors, subsectors, and the three main regions. In many cases the new series use updated weights and a two-step weighting procedure. Finally, we compare each set of results to their respective reference series, to evaluate the tracking record of responses to the activity questions. In all cases the new series are calculated in R and exported to Excel.

We have interpolated a few missing values by replacing the missing values with the published series, where available. In the cases where there were no published series we have just used simple linear interpolation to create complete continuous series. All the duplicates have been removed from the calculations. Duplications may arise due to some respondents answering in both email/fax format, as well as via post. The BER only uses the first response that it receives.

The latecomers are excluded from the calculations, as we have decided to stick to the BER's convention. The results below illustrate that the inclusion of latecomers does not influence the results substantially, with only slight changes at the subsector level. Since 2016Q2, however, the BER has included imputed values, in order to increase the number of responses and to decrease the volatility of the balance series. An imputed value refers to a latecomer in period $t - 1$, who had not responded in time in period t . For example, all the latecomers in 2017Q1, who had not responded in time in 2017Q2, were added to the 2017Q2 responses. If such a firm was also a latecomer in 2017Q2, the response may potentially be an imputed value in 2017Q3, if it is a latecomer again. The imputed values are reflected in the published results since 2016Q2, and are therefore included in the comparisons below. The balances calculated with more stable samples of firms, i.e. firms that responded in consecutive quarters or firms that responded to more than half of all surveys, were remarkably similar to the balances using the full sample. This suggests that the results are quite robust.

The report proceeds as follows: the survey results for the main sectors are evaluated separately. The report discusses the changes to the raw microdata that were necessary to produce results consistent with the BER's published series. The balances calculated from the microdata are then compared to the published series. The new series are then calculated with alternative weighting procedures. Each set of series is then compared to their respective reference series to evaluate their tracking record. Finally, the different series are compared in terms of their volatility.

2 The Building Survey

In this section we examine the Building sector surveys and results. We have used the Building survey microdata to calculate the results for the period 1993Q2-2017Q2.

2.1 Collating and Cleaning the Microdata

In order to match the questions and responses in the different versions of the surveys, the surveys in the old records were compared to the current incarnations. We identified and marked the questions in the old (pre-2001) Building surveys and then matched them to the new survey questions. The old questionnaires have been uploaded to the Dropbox folder.¹

There were a number of errors in the raw microdata files.² We corrected the errors relating to the questionnaires in the raw microdata files. There were also a number of errors relating to the region codes³, which have been corrected. For instance, there was a clear change in region codes in 1995Q1, and region code before that (1993Q2-1994Q4) differ again. We have corrected these region codes by taking the mode for each individual respondent, i.e. the codes with the highest frequencies for each ID code. The corrected Excel files have been uploaded to the Dropbox folder.

Table 1 reports some characteristics of the Building survey microdata. The Building microdata are missing for the following quarters: 1993Q4, 1998Q3, 2000Q2 and 2005Q4. Where published series were available, we used the published values for those missing values. Where they were unavailable, we used simple linear interpolation to create complete continuous series. We compared the substitution of the published values (for the series that are available) with simple linear interpolation. There is very little difference (only a few values were missing).

Table 1: Sample characteristics

	Sample	Total Obs	Obs/Quarter	Missing Quarters
Building Survey	1993Q2-2017Q2	18724	201.33	1993Q4,1998Q3,2000Q2,2005Q4

Table 2 reports the subsector codes that were used to split the sample. The subsectors that the BER calculated before are *contractors* (total, residential and non-residential), and *subcontractors* (total, residential and non-residential). The new sectors are the *residential*, *non-residential* and *total building* sectors. These aggregations are added to make the survey results comparable to the official data. Official data is only available for the residential sector, non-residential sector and the total construction sector. No high frequency series for contractors and sub-contractors are published. Figure 1 illustrates the number of respondents to the Building survey, by subsector.

¹In the scanned copies, X's denote questions that are no longer asked. Yellow denotes the Q1 consistency check - all the responses must be either 1s or 2s. Green denotes the questions that were only asked in earlier (pre-1996Q3) surveys. Red denotes apparent coding mistakes that we have corrected.

²In the pre-2001 data, Q8 and Q6 seemed to have changed, which led to large spikes (e.g. in 1995Q1). In the post-2001, responses for Q11 were recorded, but the question does not seem to exist.

³In the pre-2001 sample there were 9 region codes, but in the post-2001 sample there were 8 region codes. We assume that the following key holds in both samples: 01=WC, 05=KZN, 06=GP.

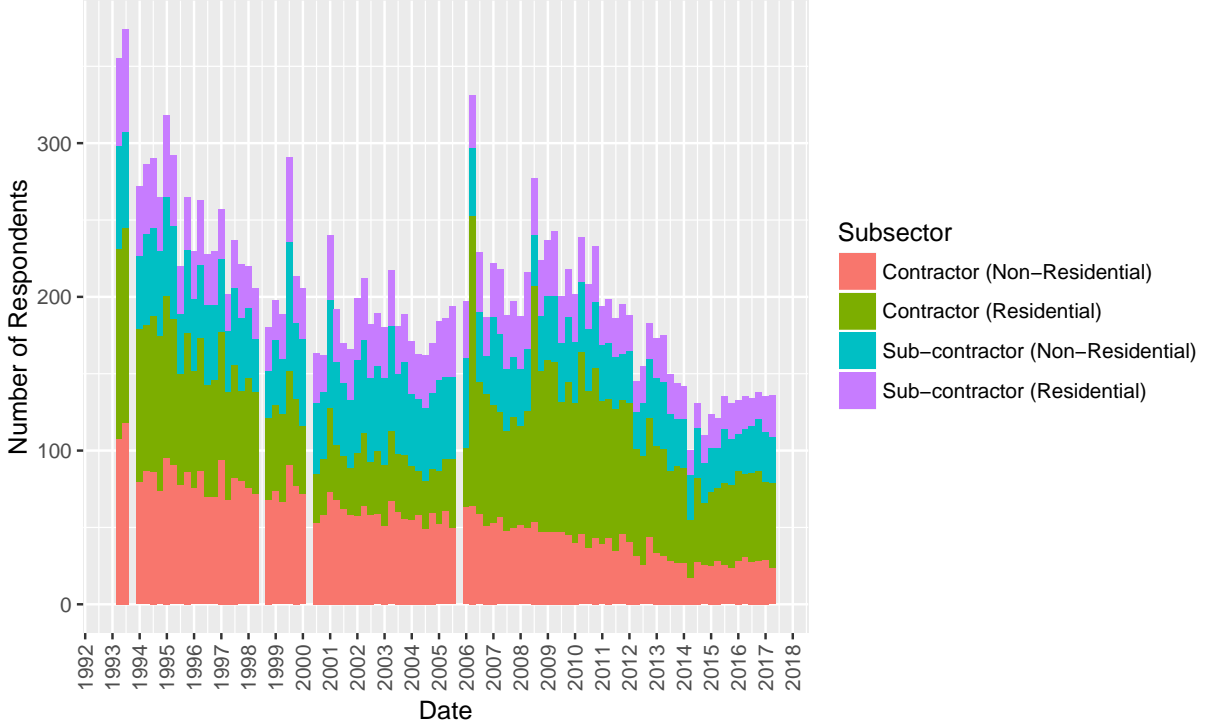


Figure 1: The number of respondents by subsector (1993Q2-2017Q2)

2.2 The Building Sector Indicators

Formally, one can define a k -period-ahead expectations measure (C_t^k) at time t as: $C_t^k = E_t f(\Delta^h Y_{t+k})$, where Y_{t+k} is a measure of real activity at time $t+k$ and $\Delta^h Y_{t+k} = Y_{t+k} - Y_{t+k-h}$. A common definition of $f(\Delta^h Y_{t+k})$ relies on an up, unchanged, or down classification (e.g. Q2A in the BER surveys):

$$f(\Delta^h Y_{t+k}) = \begin{cases} -1, & \text{if } \Delta^h Y_{t+k} < 0 \\ 0, & \text{if } \Delta^h Y_{t+k} = 0 \\ 1, & \text{if } \Delta^h Y_{t+k} > 0 \end{cases}$$

An alternative would be to use a binary classification in levels (e.g. Q1 in the BER survey):

$$f(Y_{t+k}) = \begin{cases} 0, & \text{if } Y_{t+k} < a \\ 1, & \text{if } Y_{t+k} \geq a \end{cases}$$

where a is determined by the preferences of the agent. In this case a is the subjective benchmark or threshold that determines when conditions are ‘satisfactory’, and the measure of confidence simplifies to: $C_t^k = E_t f(Y_{t+k})$.

We make a distinction between indicators of current conditions C_t^k when $k = 0$, and indicators of expected conditions C_t^k when $k = 1$. The BER business tendency surveys make this distinction possible by asking for separate responses relating to current and expected future conditions.

The series are based on balance statistics, which present a single figure summarising the responses of all participants to a particular question. It is the cross-sectional mean of survey responses if the

Table 2: Building sector codes

Sector	Code
Contractors: Residential	5000
Contractors: Non-Residential	5010
Sub-Contractors: Residential	6000
Sub-Contractors: Non-Residential	6010

standard quantification system is used: ‘better’ is quantified by +1, ‘the same’ by 0, and ‘poorer’ by -1. The balances in period t relating to current conditions C_t^0 , and the balances in period t relating to expected conditions C_t^1 , may be defined as:

$$C_t^0 = \frac{1}{W_t} \sum_{i=1}^N w_{it} E_t f(\Delta^4 Y_{i,t})$$

$$C_t^1 = \frac{1}{W_t} \sum_{i=1}^N w_{it} E_t f(\Delta^4 Y_{i,t+1}),$$

where $Y_{i,t+k}$ is a measure of real activity at time $t+k$ for firm $i = 1, \dots, N$; $\Delta^h Y_{i,t+k} = Y_{i,t+k} - Y_{i,t+k-h}$ for firm i ; w_{it} is the weight that each firm i receives at time t ; and $W_t = \sum_{i=1}^N w_i$ is the sum of the weights.

2.2.1 New Indicators

The Building survey responses do not receive weights (i.e. responses are weighted equally). The indicators (balances) for a number of groupings were calculated, as is illustrated below. A selection of the indicators are illustrated below, although balances have been calculated for all of the questions, and are included in the Excel files. Figure 2 illustrates the results for Q1 (confidence) in the three new subsector classifications. Figure 3 illustrates the balances for total construction in the Western Cape, KwaZulu-Natal and Gauteng.

2.2.2 Latecomers

This section illustrates the impact of the inclusion of the latecomers on the indicators. The latecomers are not identified in the pre-2001 sample. We compare the post-2001 results, with and without latecomers included. The results illustrate that the inclusion of latecomers does not influence the results substantially, with only slight changes at the subsector level, which is consistent with (Kershoff, 2015). Figure 4 shows that the inclusion of a large number of latecomers (10% on average) had a small effect on the survey results. Although this is not entirely unexpected at the higher levels of aggregation, it is surprising that it also had almost no effect at high levels of disaggregation. This result supports the assumption that the distribution of missing responses corresponded with that of responding firms, i.e. the so-called missing-at-random (MAR) assumption holds (European Commission, 2006). We therefore decided to stick with the convention of excluding latecomers from all of the calculations going forward, which means that the BER does not need to revise the series when they are updated.

Since 2016Q2, the BER has included imputed values, in order to increase the number of responses and to decrease the volatility of the balances. An imputed value refers to a latecomer in period

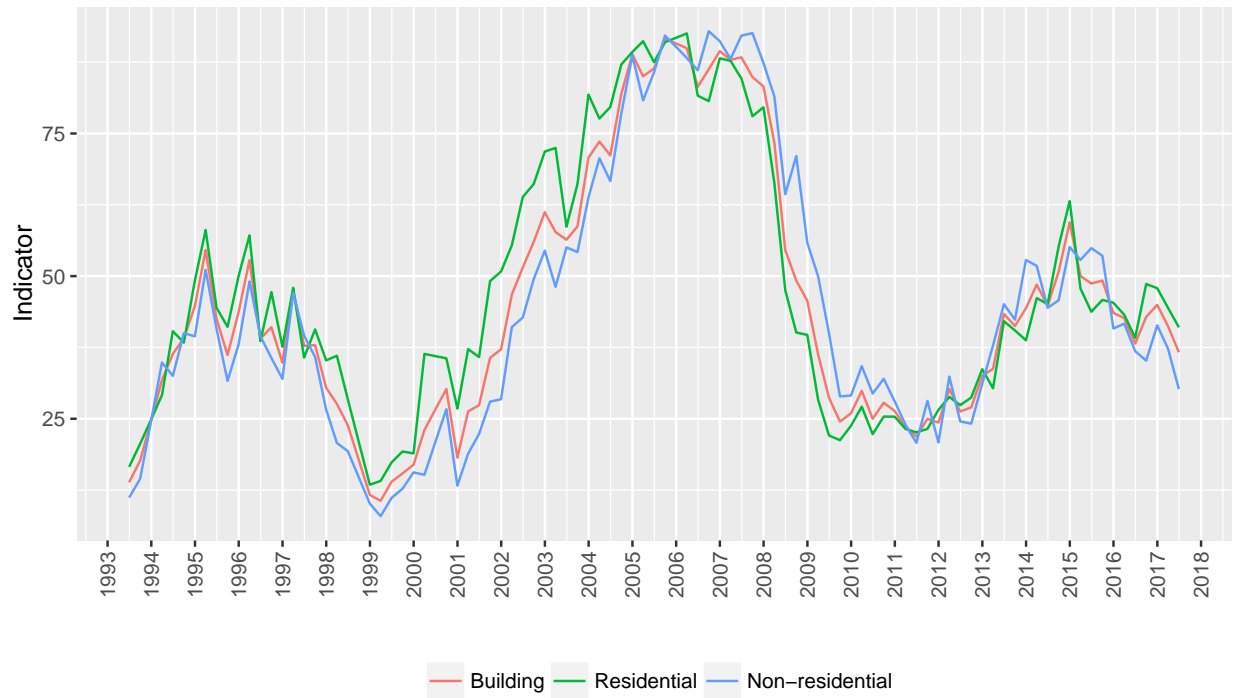


Figure 2: Building results for Q1 (confidence) in the new categories

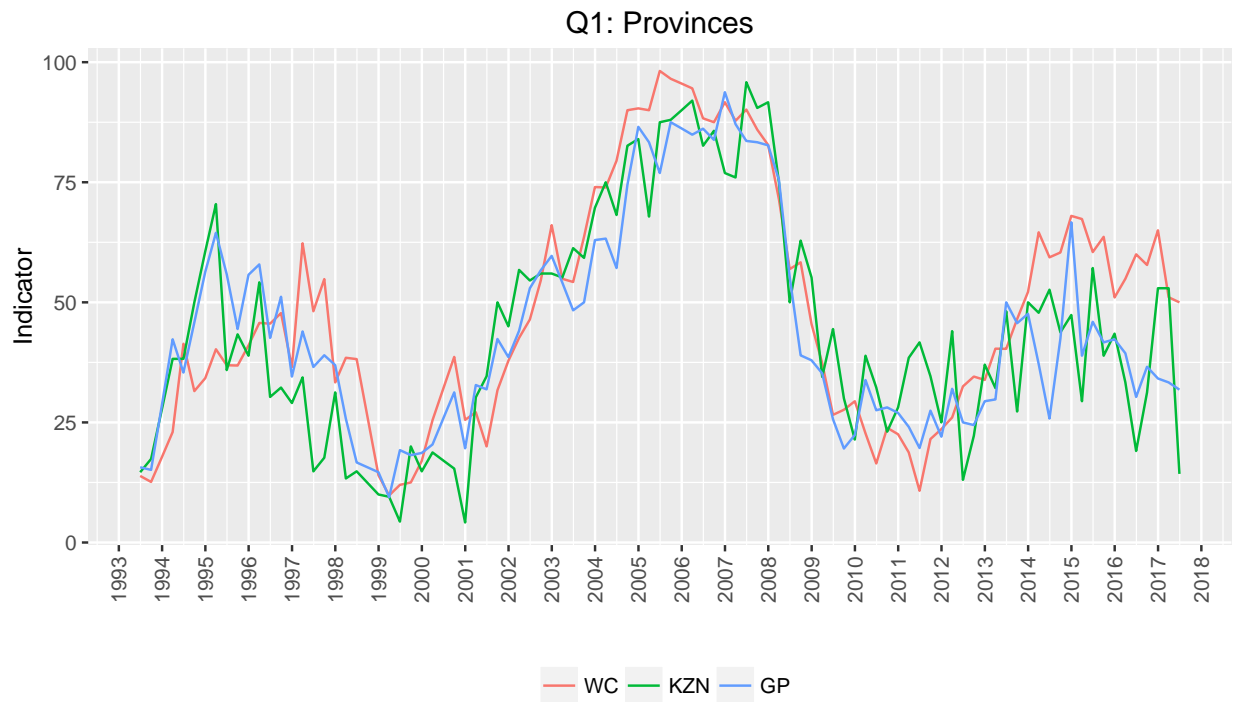


Figure 3: Building results for Q1 (confidence) for the main provinces

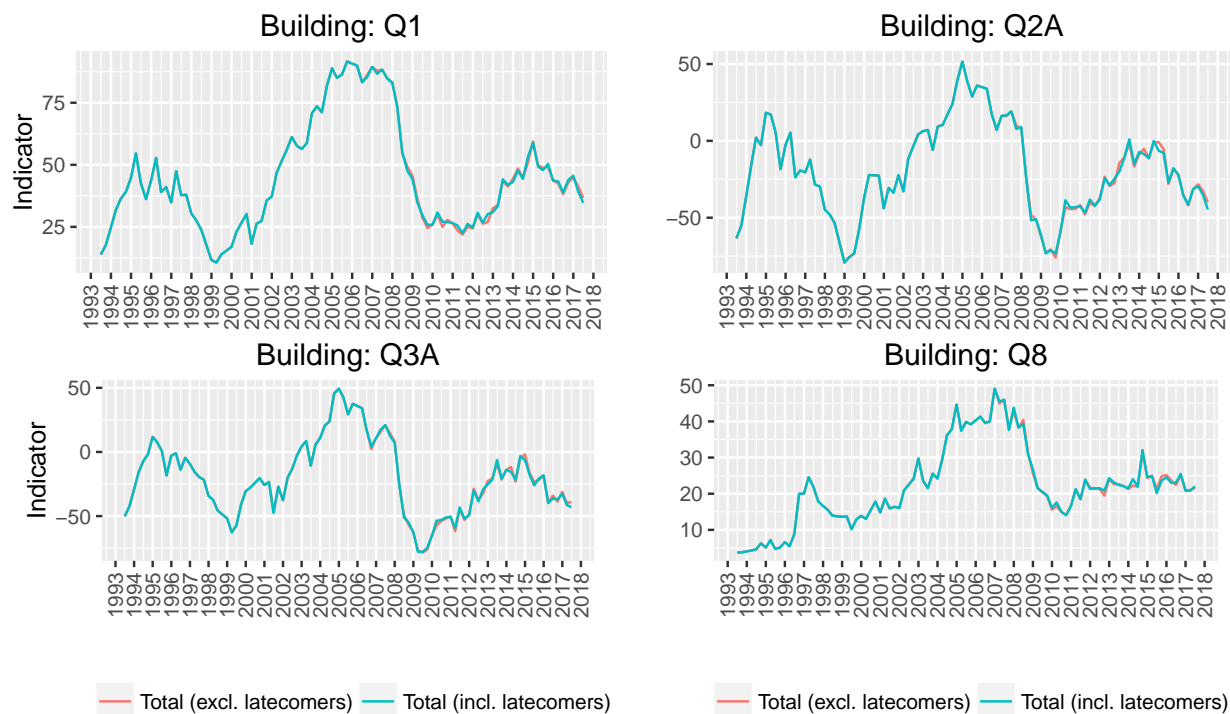


Figure 4: Building results for Q1 (confidence) - including and excluding latecomers

$t - 1$, who had not responded in time in period t . For example, all the latecomers in 2017Q1, who had not responded in time in 2017Q2, were added to the 2017Q2 responses. If such a firm was also a latecomer in 2017Q2, the response may potentially be an imputed value in 2017Q3, if it is a latecomer again. The imputed values are reflected in the published results after 2016Q2, and are therefore included in the comparisons below. In future research we can evaluate the impact of this inclusion going back to 2001.

2.3 Published Series

In order to check that the microdata and the calculations are correct, we compare the estimated indicators using the microdata to the BER's published series, where available (the "BERdata_Build.xlsx" file on Dropbox).⁴ Figure 5 and Figure 6 compare the series calculated with the microdata to the BER's published series. The results are very similar, albeit with a few small discrepancies. In addition, the microdata provides longer series for the questions on the constraints than the published series. The microdata therefore seems to be adequate for calculating the new balance statistics for all the questions, as well as for the updates.

⁴The "published" series for the constraints (Q7-Q9) in the Excel file "BERdata_Build.xlsx" have a few strange values, but these are not published on Quantec. We decided to exclude these values.

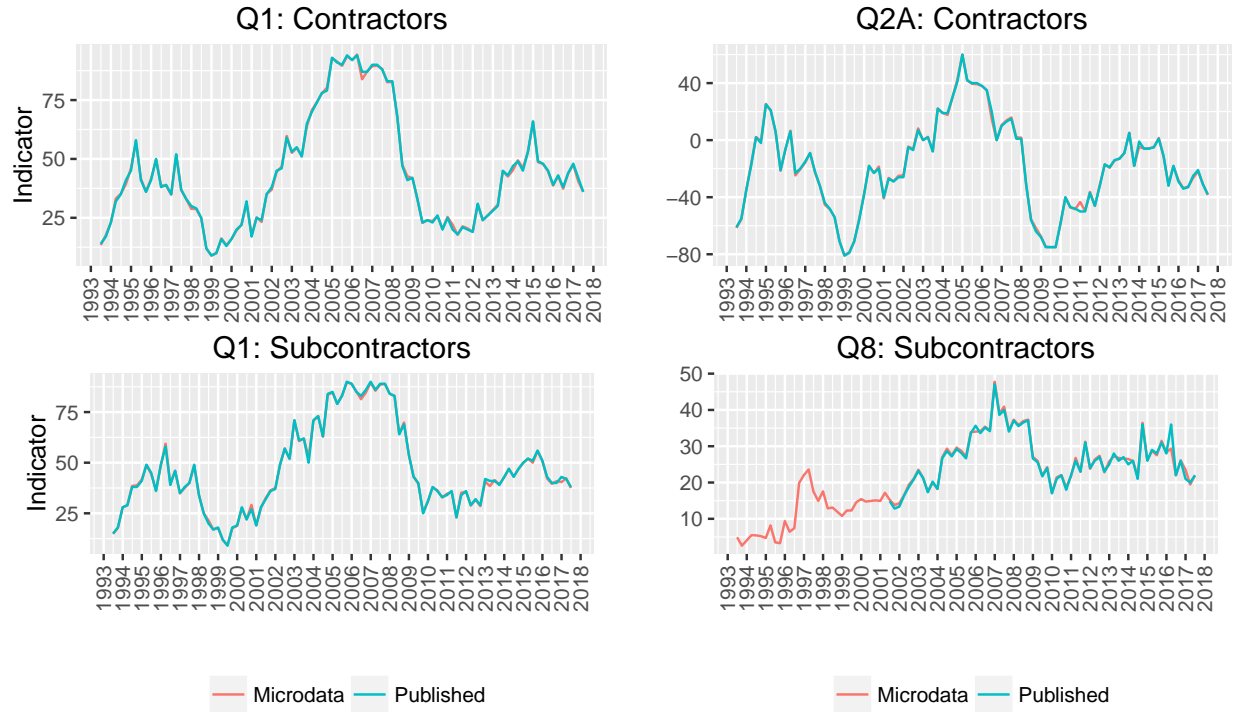


Figure 5: Microdata results compared to the published results

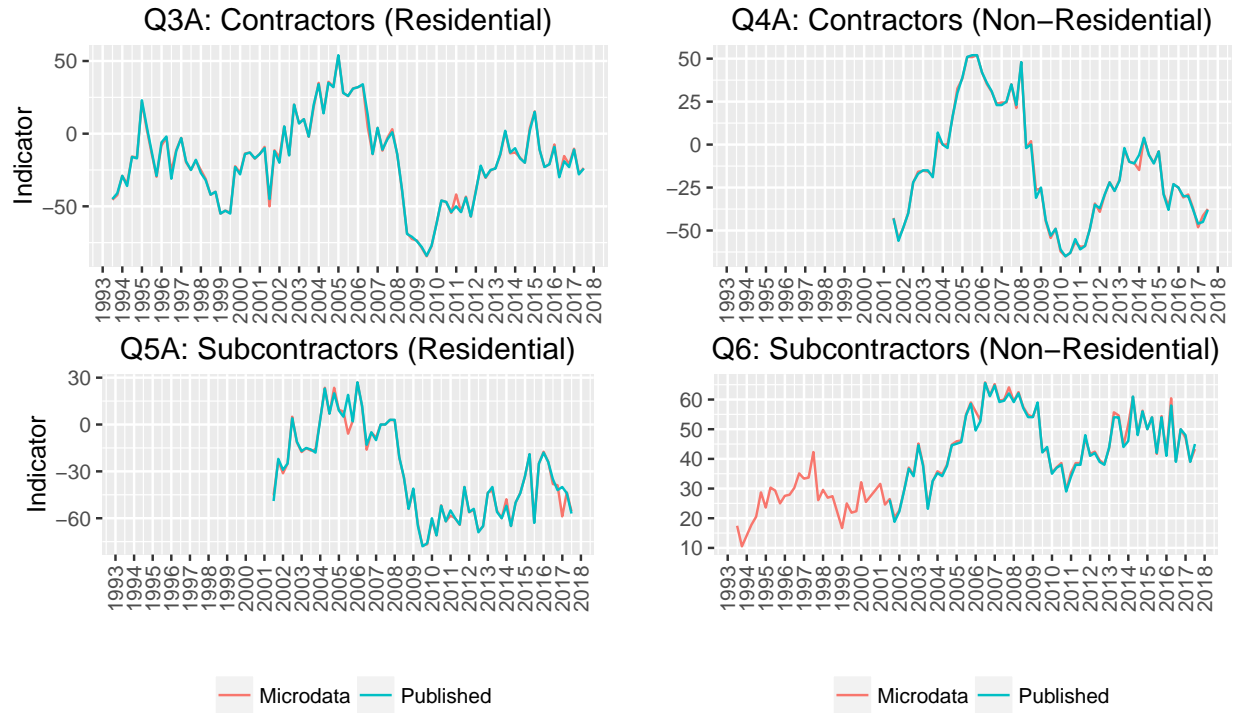


Figure 6: Microdata results compared to the published results

2.4 Alternative Weights

The balances for the Building surveys are calculated without applying weights (i.e. an equal weighting procedure). The results are similar when the firm size weights are used. In order to be consistent throughout all the BER surveys, I would suggest that in future research we should create a weighted version of the Building indicators. We would need to calculate subsector size weights and the firm size weights should also be converted to four categories, as we have done with the other surveys. Once we have these weights we can apply the 2-step weighting procedure.

2.5 Reference Series

In this section the series are assessed in terms of their tracking record, i.e. their correlations to their respective reference series. We compare the series graphically and calculated simple correlations with alternative potential reference series from official statistics.

The question is which reference series should be used. The SA Reserve Bank (SARB) produces quarterly data on residential and non-residential fixed investment, while Stats SA produces monthly data on building plans passed and completed. These data all have their respective advantages and disadvantages. The SARB's data is the best match for the BER's survey, as it provides for the spread of activity over time (using an S-curve for adjustments). However, its disadvantage is that the recent data is revised extensively when more information becomes available and no provincial data is available. Stats SA published a series for building plans passed, which is a good leading indicator, but postponements and the stock of plans makes this data problematic as a reference series (or benchmark) for the BER's building survey. Building plans passed would not necessarily have a stable relationship with activity, as plans that are approved are not necessarily completed. Building plans completed are useful, bearing in mind that Stats SA only records the activity upon completion, while the BER's survey records the progress from start to completion.

A comparison of the SARB and Stats SA data shows that they are closely correlated, even though Stats SA record activity only at the end of the period, while the SARB spreads it over the whole period of construction. The comparison also shows that the Stats SA data is more volatile, especially for non-residential activity. This is to be expected, given the lumpy nature of non-residential construction.

Stats SA publishes data for residential buildings, non-residential buildings, additions and alterations, and total activity. To make this data comparable to those of the SARB and the BER, additions and alterations were divided between residential and non-residential buildings. As a final step, other residential (e.g. hotels and casinos) and non-residential buildings (e.g. schools, sport clubs, churches and hospitals) were excluded to make the data even more comparable to the BER's, as such buildings may have a big impact on the growth rate in specific periods and the BER does not always cover them effectively. However, the "other" category has little impact on the series over longer time periods.

In this case we compare the series for general business conditions (Q2A) and total activity (Q3A) to the following reference series:

- YoY% change in the SARB's real residential & non-residential gross fixed capital formation (SARB GFCF)
- YoY% change in StatsSA's real building plans completed;

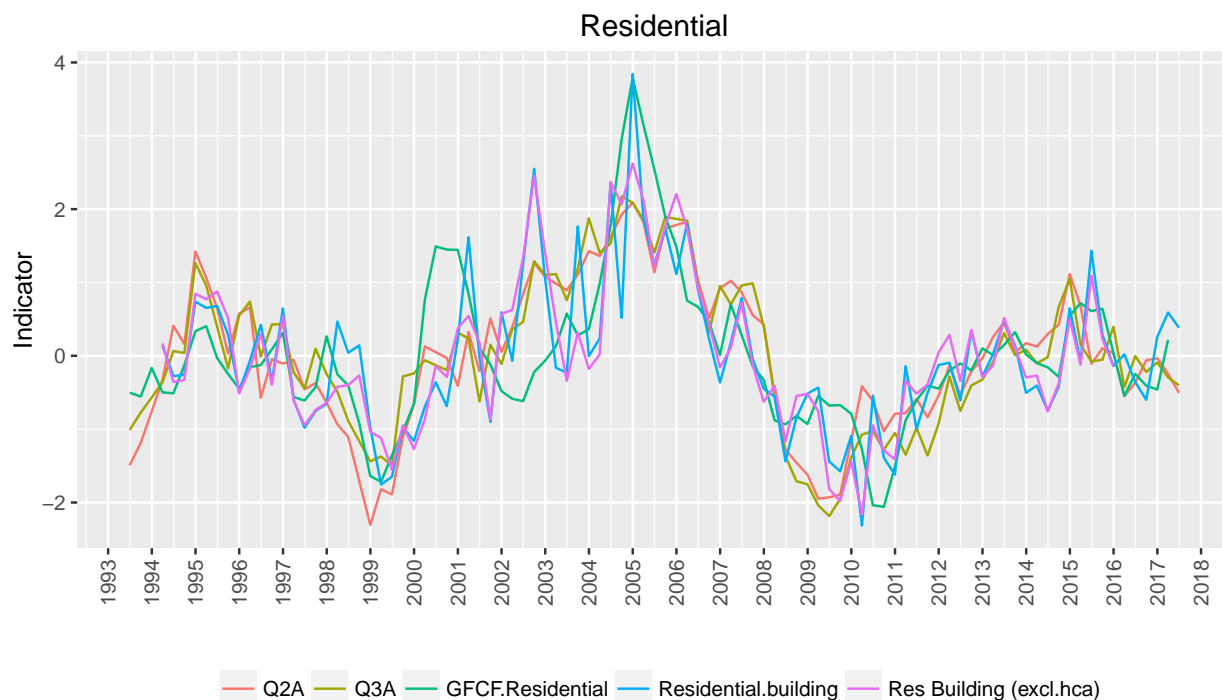


Figure 7: Residential results compared to the reference series

- YoY% change in StatsSA’s real building plans completed, excluding big hotels, casinos and alterations (“additions and alterations” do not require building plans).

2.5.1 Residential Activity

Figure 7 illustrates the Residential series compared to the SARB and Stats SA references series. Table 3 reports that the correlations between the survey balances and the reference series are relatively high. This is confirmed by the cross-correlations illustrated in Figure 8. These show that the reference series with the highest correlations are Stats SA’s building plans completed (excluding big hotels, casinos and alterations).

Table 3: Residential series correlations		
	Q2A	Q3A
GFCF.Residential	0.72***	0.73***
Residential.building	0.70***	0.71***
Res Building (excl.hca)	0.77***	0.77***

2.5.2 Non-Residential Activity

Figure 9 illustrates the Non-Residential series compared to the SARB and Stats SA references series. In this case the reference series are very volatile. Table 4 reports that the correlations between the survey balances and the reference series are relatively low, although the cross-correlations in Figure

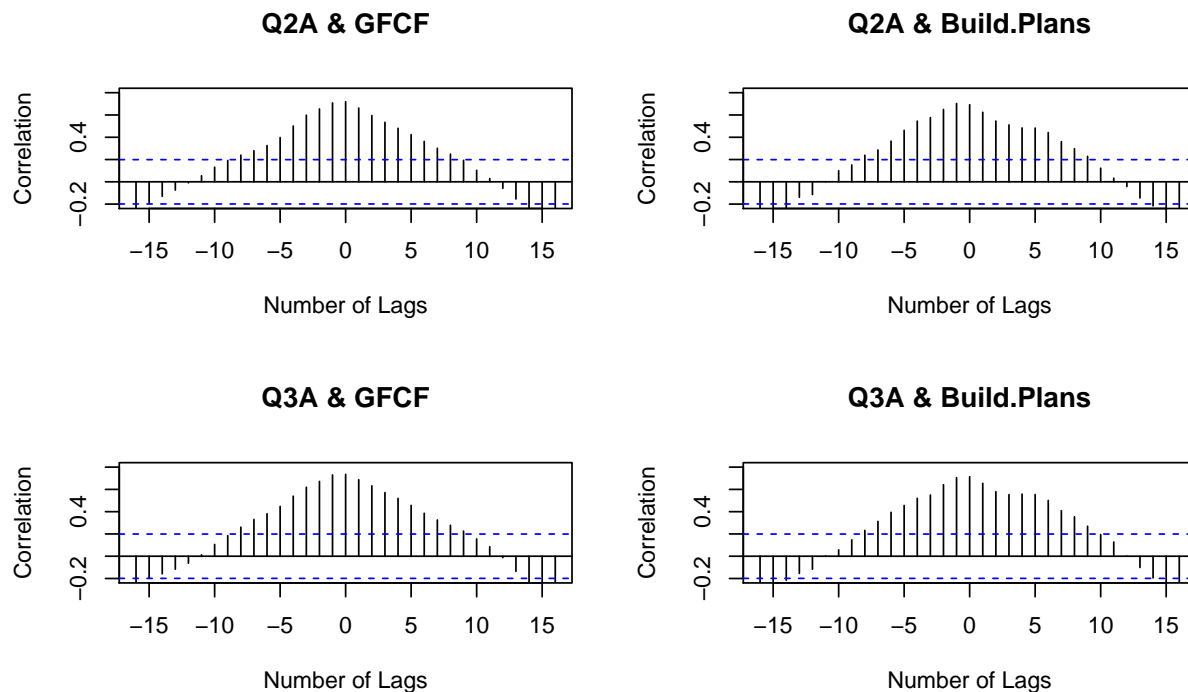


Figure 8: Residential series cross-correlations

10 show that the correlations are higher when the balances are lagged, i.e. the survey results have a leading relationship with the reference series. The highest correlations with the SARB series are when they lead by around 2 quarters and with the Stats SA series when they lead by around 4 quarters.

Table 4: Non-residential series correlations

	Q2A	Q3A
GFCF.Non-residential	0.46***	0.45***
Non.residential.building	0.27***	0.33***
Non-Res Building (excl.hca)	0.34***	0.40***

2.5.3 Total Building Activity

Figure 11 illustrates the Total Building Activity series compared to the SARB and Stats SA references series. Table 5 reports that the contemporaneous correlations between the survey balances and the reference series are relatively high. The correlations of the survey balances with the residential plus non-residential GFCF series are much higher than the correlation with the total GFCF series. The high correlations are confirmed by the cross-correlations illustrated in Figure 12, which show again that the survey balances lead the Stats SA reference series (building plans completed) by a few quarters.

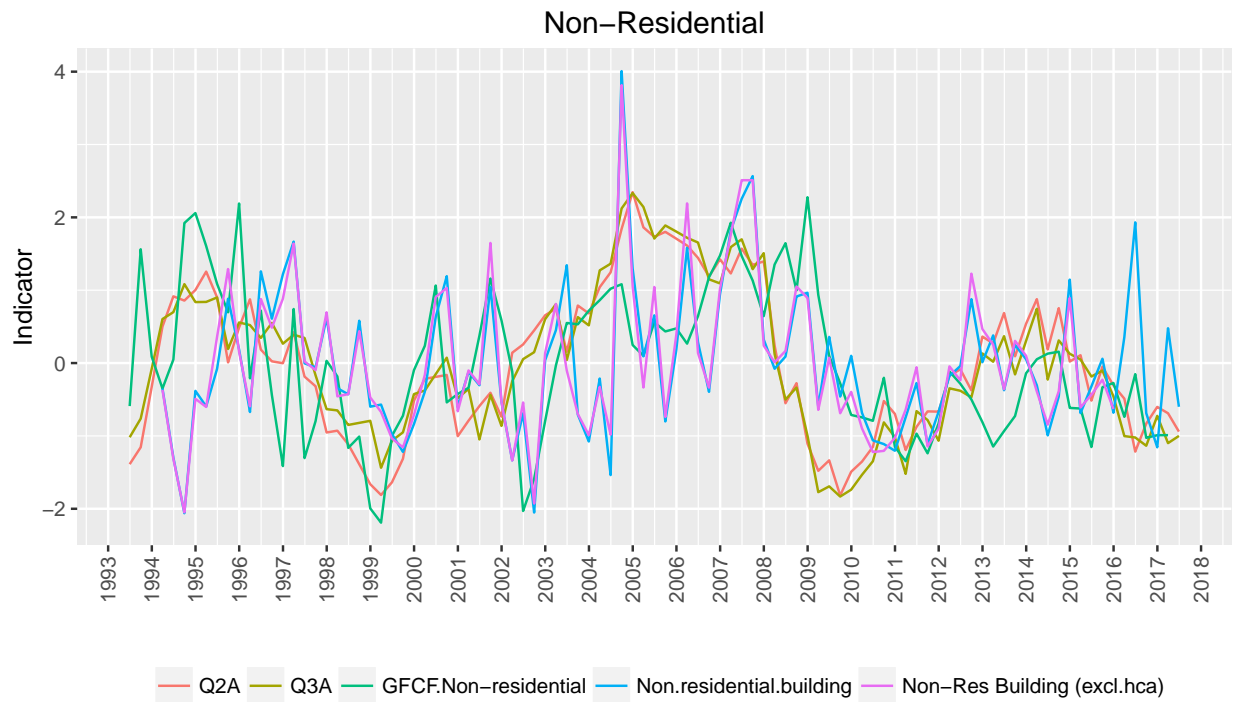


Figure 9: Non-residential results compared to the reference series

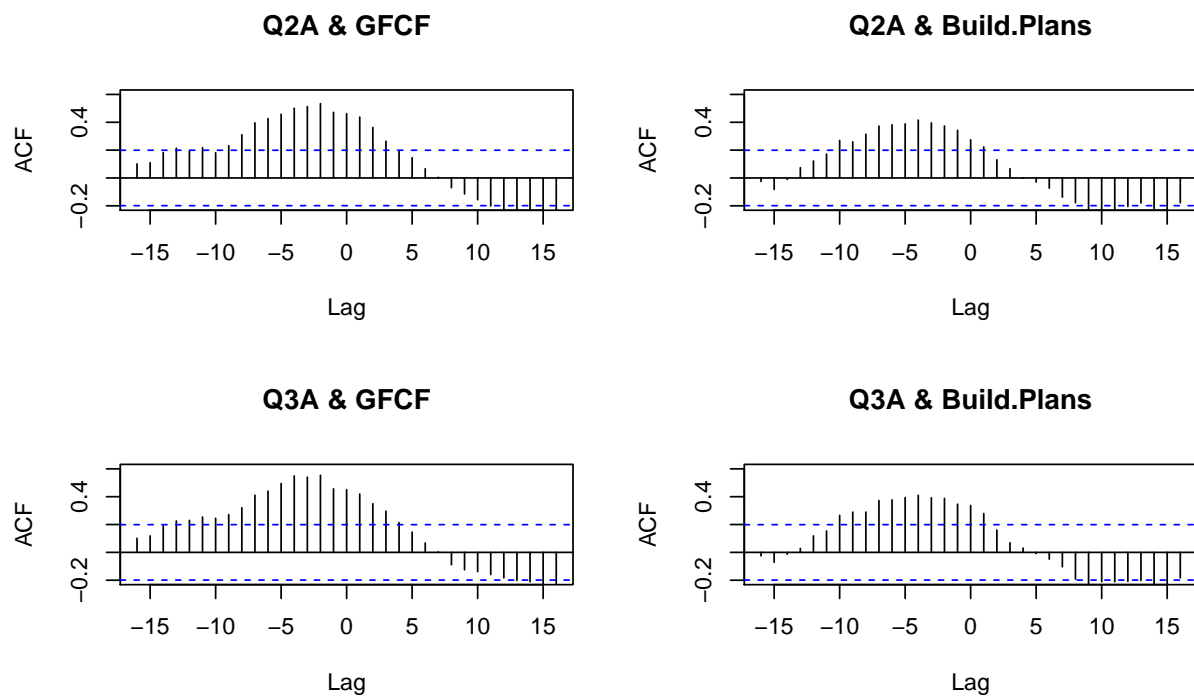


Figure 10: Non-residential series cross-correlations

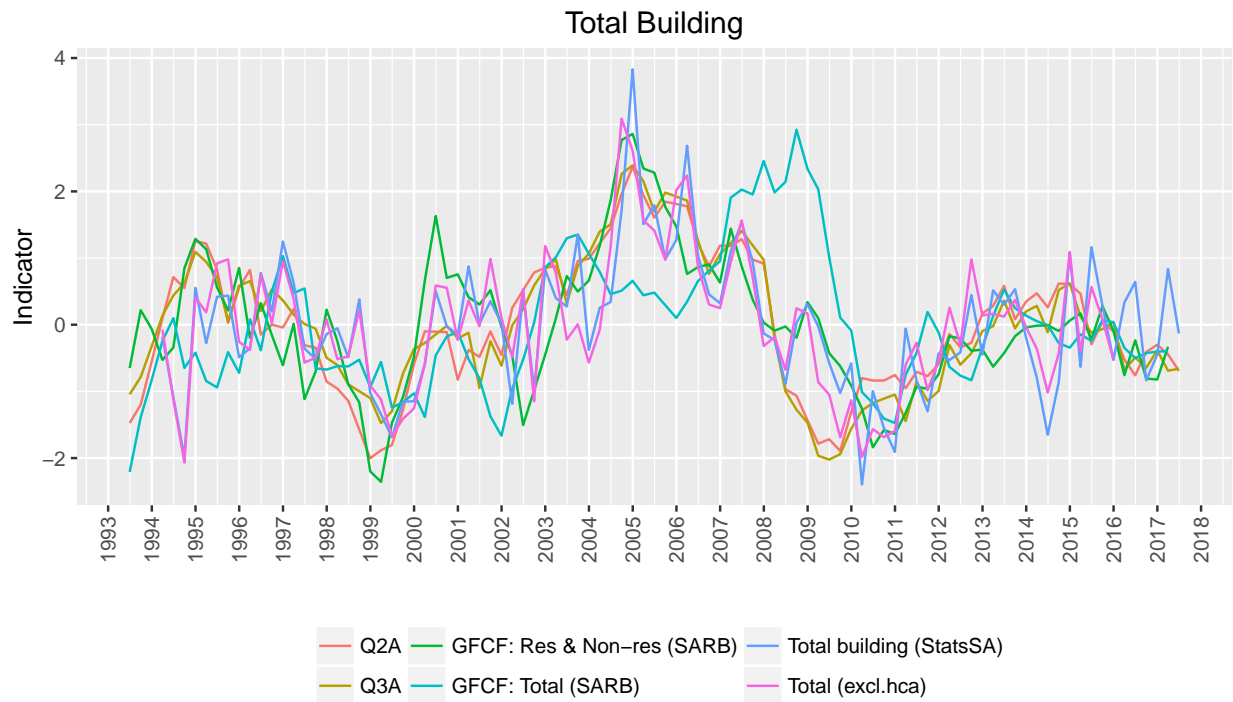


Figure 11: Total results compared to the reference series

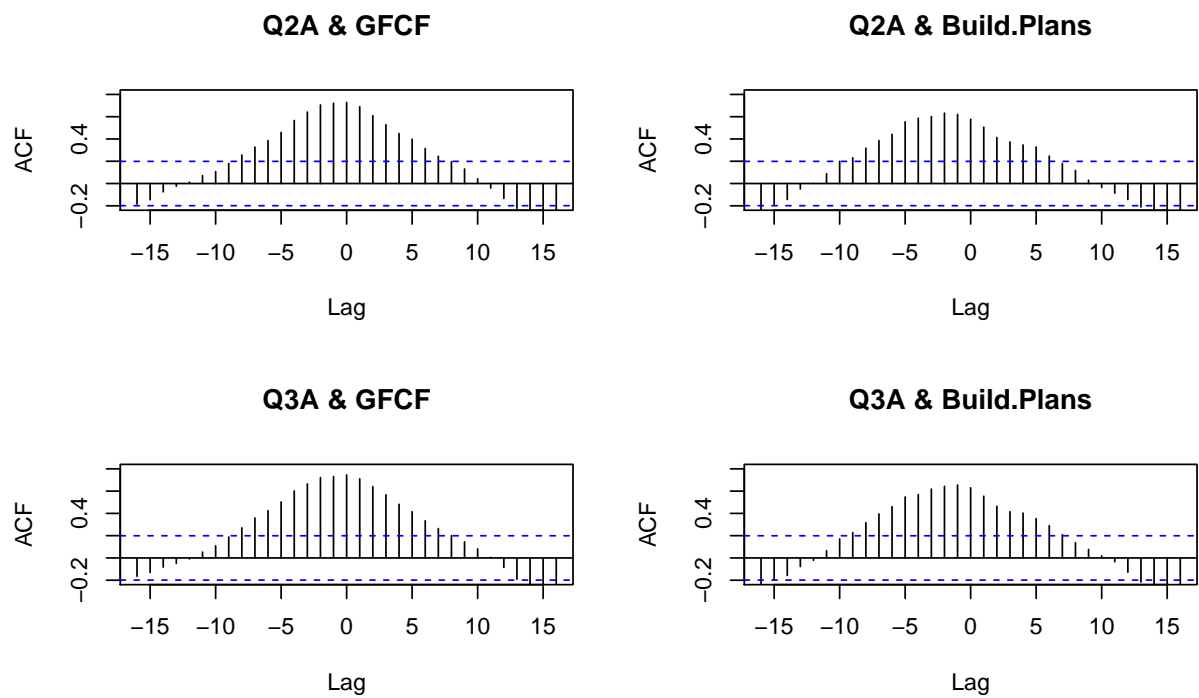


Figure 12: Total series cross-correlations

Table 5: Total series correlations		
	Q2A	Q3A
GFCF: Res & Non-res (SARB)	0.73***	0.74***
GFCF: Total (SARB)	0.31***	0.30***
Total building (StatsSA)	0.58***	0.63***
Total (excl.hca)	0.66***	0.70***

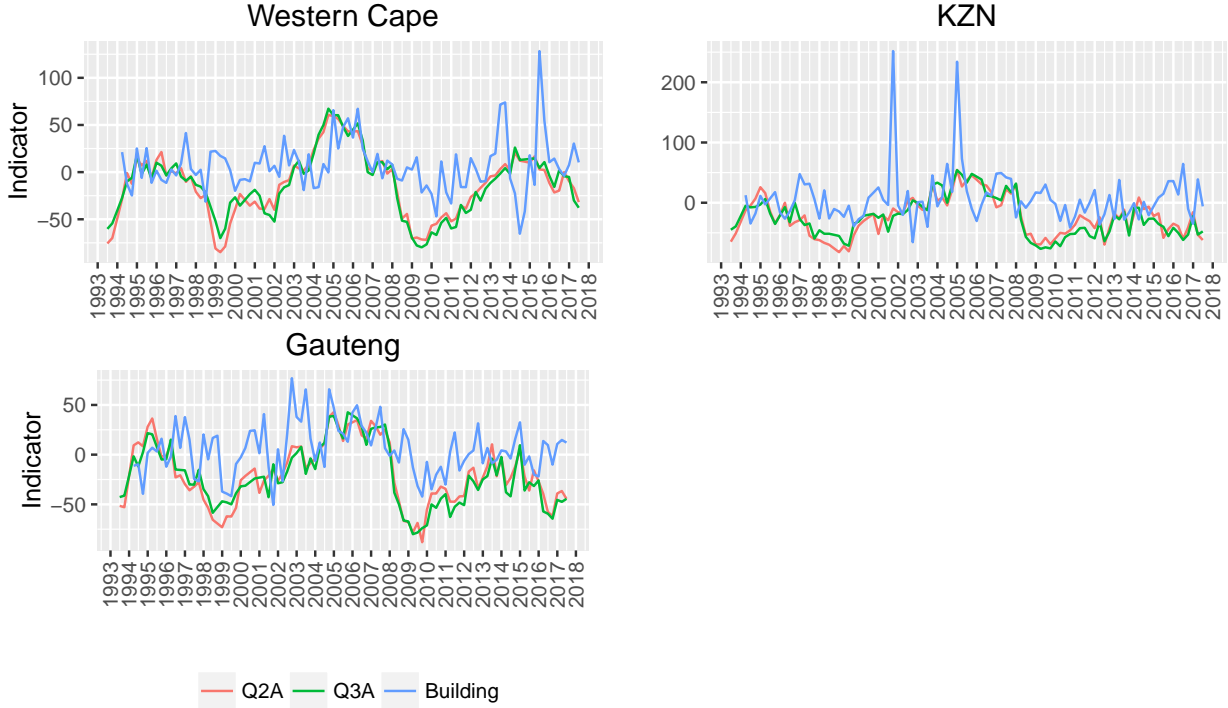


Figure 13: Regional results compared to the reference series

2.5.4 Regional Activity

Figure 13 illustrates the Regional series compared to the Stats SA references series (the only one that is available). In this case the reference series are extremely volatile. Table 6 reports that the correlations between the survey balances and the reference series are relatively low due to this volatility, although they are marginally higher when the survey balance series are lagged, illustrated in Figure 14.

For simplicity, and because the disaggregated data are available by province, we decided to use the simple Stats SA building plans completed as reference series. This means that we do not make adjustments for big projects, such as hotels and casinos, or alterations, even though such an adjustment increases the tracking record of the series somewhat. The Stats SA series is also the one that is published by province, which simplifies the comparison and makes it consistent for all the series. For the remainder of the document, therefore, where possible we focus on the Stats SA reference series.

This section has demonstrated that the Building survey microdata are consistent with the published series. The survey balances seem to track the reference series relatively well. In future research we

Table 6: Regional series correlations

	Building
WC.Q2A	0.27***
WC.Q3A	0.30***
KZN.Q2A	0.21**
KZN.Q3A	0.21**
GP.Q2A	0.43***
GP.Q3A	0.44***

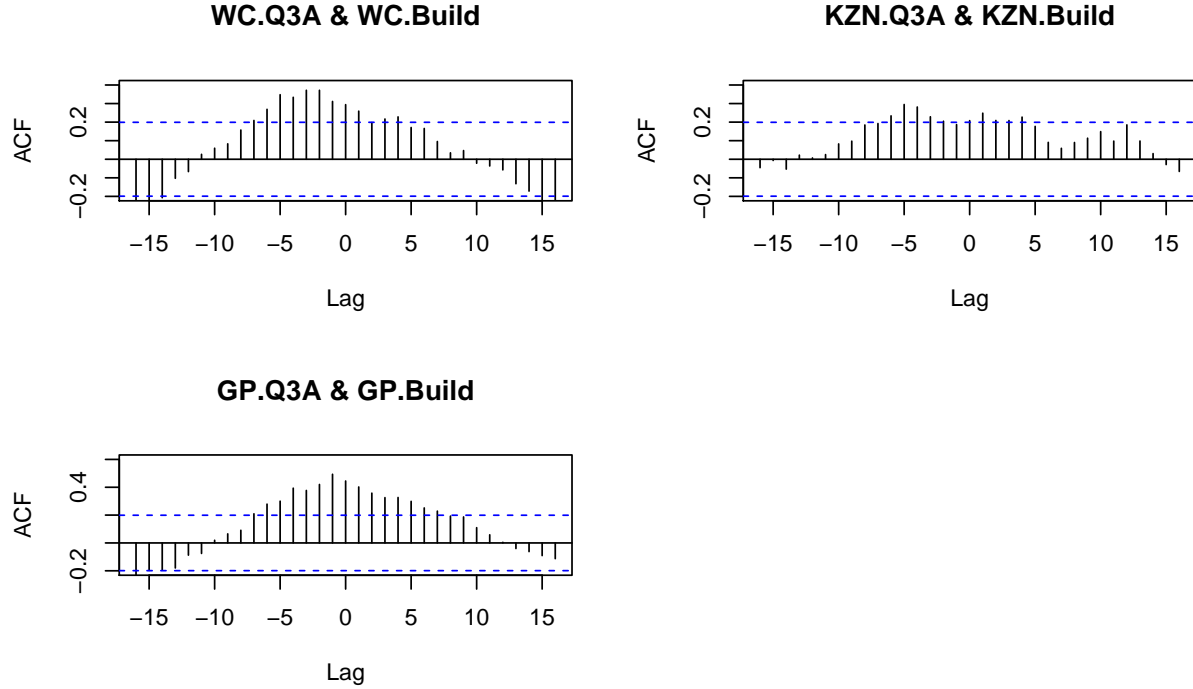


Figure 14: Regional series cross-correlations

should calculate weighted versions of the Building balance series, in order to make all the calculations consistent.

2.6 Architects, Quantity Surveyors and Civil Construction

The microdata for the surveys of architects, quantity surveyors and civil engineers is only available from 2001 (Argitekte_01Q2-17Q2.xlsx, Civils_01Q2-17Q2.xlsx and Quantity_01Q2-17Q2.xlsx). These series do not form part of the overall business confidence indicator. As with the Building survey, the responses do not receive weights (equal weighting procedure), even though firm size weights are available. Figure 15 compares the results from the microdata to the published series. They are very similar and the microdata seem to be correct. In future we should consider incorporating (weighted versions) of these three surveys into the Building sector results.

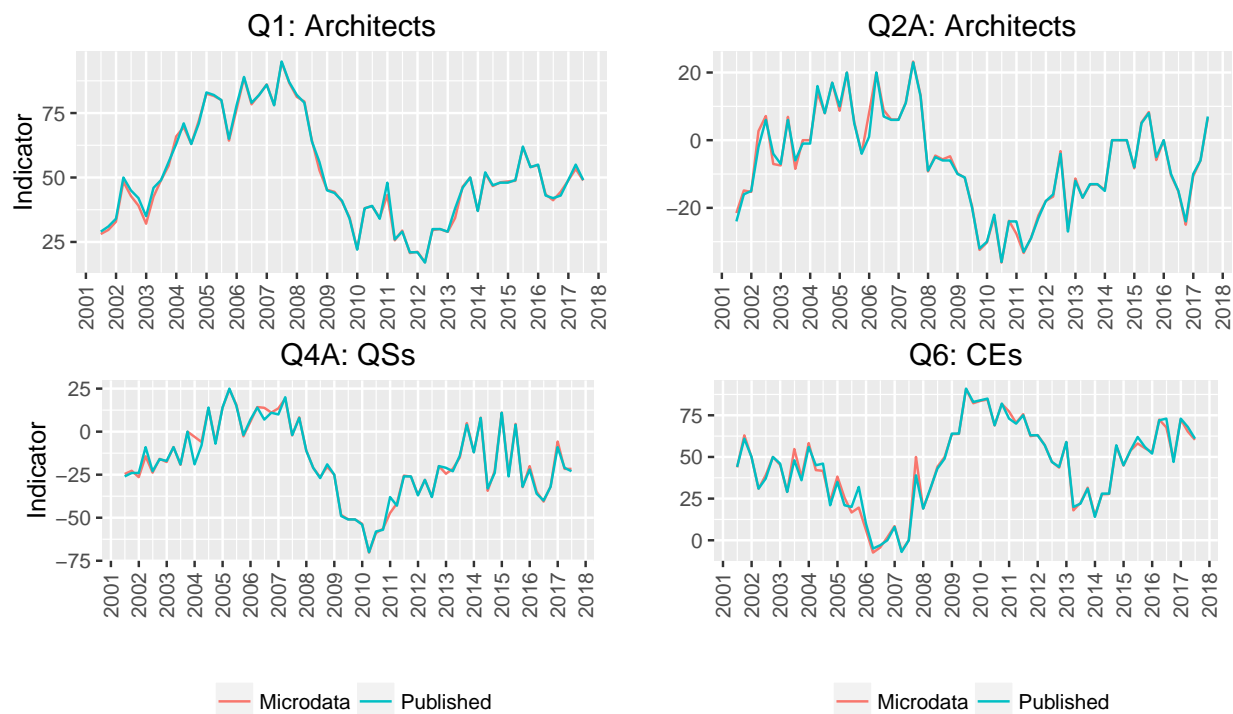


Figure 15: Architect, QS and Civil Engineer results compared to the published results

2.7 Summary

After some corrections the microdata seems to provide series that correspond very closely to the published series. We have decided to exclude latecomers to make updating easier and avoid revisions, because they make a small difference anyway. We have calculated new series and compared them to the reference series. We chose the Stats SA series as reference series, as it is widely available for many of the subsamples and has a relatively high correlation with the balances. The final series are written to an Excel file. The updates can and should be fully automated.

3 The Manufacturing Survey

In this section we examine the Manufacturing sector surveys and results. We have used the Manufacturing survey microdata to update the results for the period 1992Q1-2017Q2.

3.1 Collating and Cleaning the Microdata

Similar to the Building Survey, in order to match the questions and responses in the different versions of the surveys, the surveys in the old records were compared to the current incarnations. We identified and marked the questions in the old (pre-2001) Manufacturing surveys and then matched them to the new survey questions. The old questionnaires have been uploaded to the Dropbox folder.

There were many clear mistakes in the sector and regions codes in the microdata, especially for the period 2000Q4-2001Q1. To correct the region codes we used the codes with the highest frequencies for each ID around that period (i.e. the mode for each individual respondent).

Table 7 reports some characteristics of the Manufacturing survey microdata. The Manufacturing microdata are missing for the following quarters: 1997Q4, 2000Q1 and 2005Q4. Where published series were available, we used the published values for those missing values. Where they were unavailable, we used simple linear interpolation to create complete continuous series. There is very little difference between the substitution of the published values and simple linear interpolation.

Table 7: Sample characteristics

	Sample	Total Obs	Obs/Quarter	Missing Quarters
Manufacturing Survey	1992Q1-2017Q2	35389	364.84	1997Q4,2000Q1,2005Q4

Table 8 reports the new subsector codes that are used to calculate balances. The aggregation is increased by decreasing the number of subsectors from 21 (22 including petroleum) to 10. Figure 16 illustrates the number of respondents by subsector, according to the new sector classifications. In addition, three new overarching subsector series are calculated for consumer, intermediate and capital goods, reported in Table 9. Many of the original sector codes were not included in the new sector classifications and these were reclassified to correspond to the new sector code classifications.

3.2 The Manufacturing Sector Indicators

As discussed above, the balances are calculated as:

$$C_t^0 = \frac{1}{W_t} \sum_{i=1}^N w_{it} E_t f(\Delta^4 Y_{i,t})$$

$$C_t^1 = \frac{1}{W_t} \sum_{i=1}^N w_{it} E_t f(\Delta^4 Y_{i,t+1}),$$

where $Y_{i,t+k}$ is again a measure of real activity at time $t+k$ for firm $i = 1, \dots, N$; $\Delta^h Y_{i,t+k} = Y_{i,t+k} - Y_{i,t+k-h}$ for firm i ; w_{it} is the weight that each firm i receives at time t ; and $W_t = \sum_{i=1}^N w_{it}$ is the sum of the weights.

Table 8: Manufacturing subsector weights

New BER classification	Old BER classification	Sectors
Food & beverages	Food	1010, 1011, 1013, 1019
	Beverages	1020, 1021
Textiles, clothing, leather & footwear	Leather	1120
	Textiles	1040, 1049
	Clothing	1042, 1060
	Footwear	1070
Wood, paper, printing and publishing	Paper & prod	1109
	Print & publishing	1110
	Wood	1080, 1081
Petroleum, chemical products, rubber & plastic	Rubber	1130
	Chemicals	1140, 1149
	Plastics	1219
	Petroleum	Missing (1220)
Glass and non-metallic mineral products	Non-metal minerals	1153, 1159
Basic metal, metal products and machinery	Basic metals	1160, 1161
	Metal products	1170, 1179
	Machinery (excl. electrical machinery)	1181, 1182, 1189
Electrical machinery	Electrical machinery	1190, 1191, 1199
Radio, TV and professional equipment		1192, 1194
Motor vehicles, parts and accessories & transport equipment	Transport equipment	1200, 1201, 1209
Furniture and other, incl. tobacco	Furniture	1090, 1099

Table 9: Main industrial groupings

Main Industrial Group	BER code	Description
Consumer Goods	1010	Meat, fish, fruit, vegetables, oils
	1011	Dairy products
	1019	Other food
	1049	Other textiles
	1060	Wearing apparel & articles of fur
	1070	Footwear
	1090	Furniture
	1099	Other (incl. tobacco)
	1110	Printing & reproduction of recorded media
	1120	Leather & leather products
	1149	Other chemical products
	1189	Domestic appliances
	1192	Radio, TV & communication apparatus
	1020, 1021	Beverages
Intermediate Goods	1013	Grain mill products
	1040	Spinning, weaving, yarns
	1042	Knitted & crocheted articles
	1080	Wood & wood products
	1081	Sawmilling
	1109	Paper & paper products
	1130	Rubber
	1140	Chemical products
	1153	Glass
	1159	Other non-metallic mineral products
	1160	Basic iron & steel & castings thereof
	1161	Basic precious & non-ferrous metal products
	1179	Other fabricated metal products
	1191	Electricity distribution apparatus
	1199	Batteries
Capital Goods	1219	Plastic
	1170	Structural metal products
	1181	Special purpose machinery
	1182	General purpose machinery
	1189	Office machinery, computers
	1190	Electrical motors, generators, transformers
	1194	Medical appliances, photographic equipment, watches
	1200	Motor cars
	1201	Parts & accessories for motor vehicles
	1209	Other transport equipment

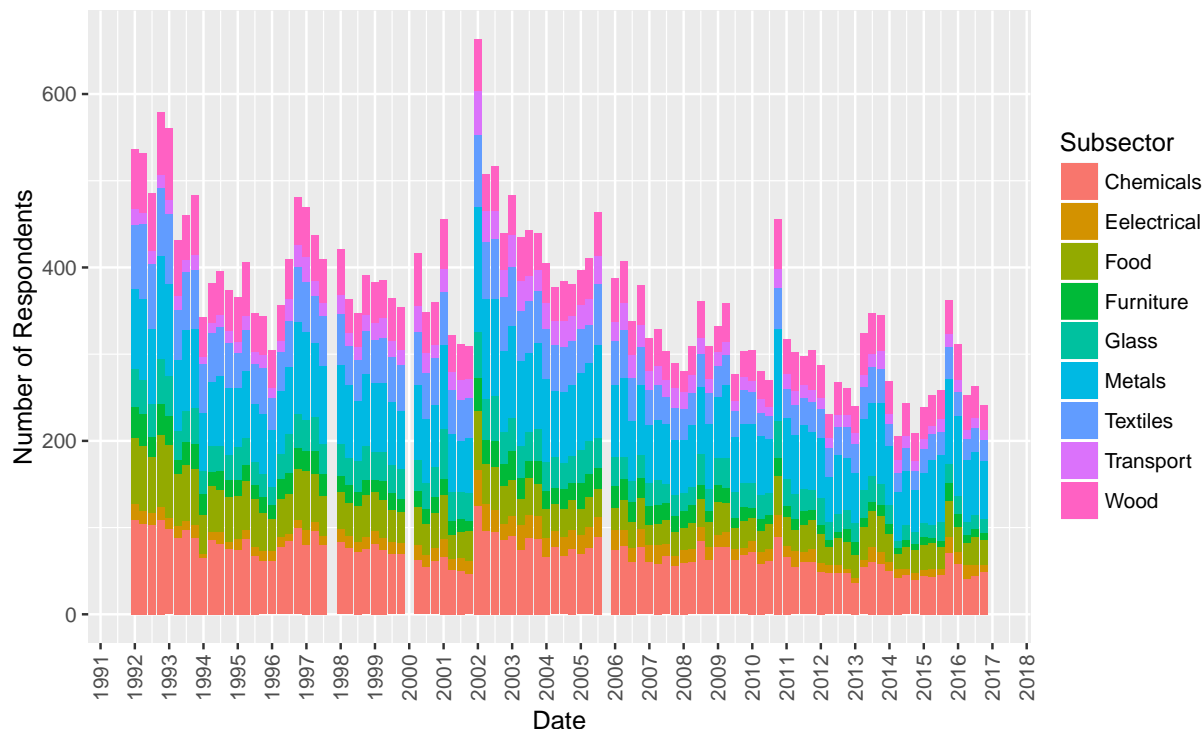


Figure 16: The number of respondents in the manufacturing sector by subsector (1992Q1-2017Q2)

For the Manufacturing survey we have created four versions of the indicators - i.e. the balance statistics for all of the questions. These balances only differ in their weighting procedure, as follows:

The baseline balance statistics for all of the questions in the Manufacturing sector are calculated in the same way as the normal BER procedure. The weights are calculated as: $w_{it} = f_{it}s_{jt}$, where f_{it} the firm size weight (i.e. the inner weight reflecting turnover or number of employees) for firm i at time t ; and s_{jt} is the subsector weight (i.e. the outer weight reflecting the share of total value added) for subsector j at time t . The firm size weights are divided into nine categories, based on the size of the workforce, and each corresponds to an exponentially increasing weight (from 1 to 700). The subsector size weights are updated periodically by the BER, based on the turnover or the number of employees in each subsector. These series are then compared to the BER's published series to ensure that the microdata is correct.

The second set of indicators are based on unweighted means, i.e. the responses are weighted equally. The third set of indicators is based on a new 4-part firm classification, which are less exponential than the previous weights. The four firm size weight categories are based on Stats SA's and the DTI's categories and thresholds of micro, small, medium and large enterprises and the classification is based on turnover rather than the size of the workforce. The idea is to see if the weighting procedure imparts unnecessary increased volatility into the series.

The final set of indicators is based on a 2-step weighting procedure, recommended by the OECD and the UN Handbooks (Organisation for Economic Co-operation and Development, 2003; United Nations, 2015). In this version the weights are calculated as: $w_{it} = f_{it}s_{jt}/F_{jt}$, where f_{it} the firm size weight (i.e. the new 4-part classification) for firm i at time t ; s_{jt} is the subsector weight for subsector j at time t ; and $F_{jt} = \sum_{i=1}^N f_{it}$ is the total firm weight for subsector j at time t . In other

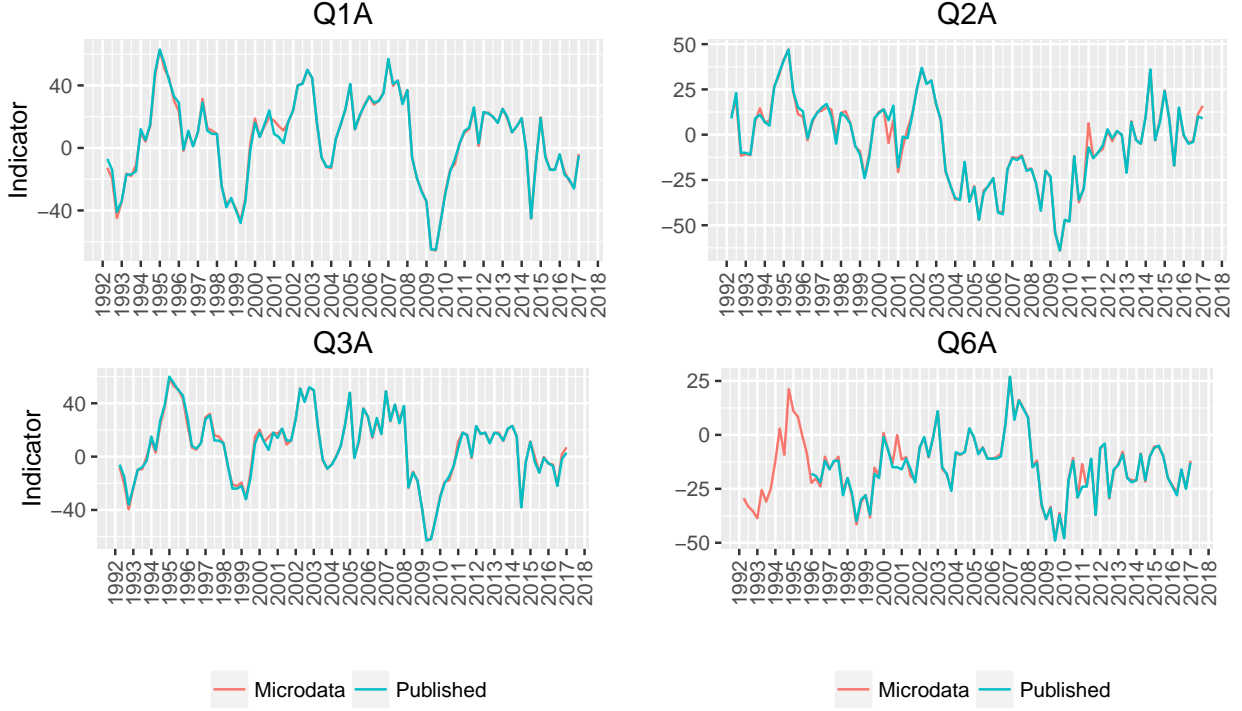


Figure 17: Manufacturing microdata results compared to the published results

words, the only difference is that we divide by F_{jt} . These weights are equivalent to an explicit 2-step weighting procedure, whereby weighted means are calculated for each subsector separately (using firm size weights), and then aggregated with the subsector weightings (United Nations, 2015).

3.2.1 Published Series

The first step is to check that the series calculated with the microdata corresponds to the BER's published series. Figure 17 compares the microdata series to the published series for a number of questions. In some cases, the microdata provides longer series than the published series. Figure 18 compares the microdata series to the published series, specially relating to the questions on constraints. The microdata series are slightly higher than the published results (although the correlations are nearly unity), implying that some adjustment factors were added to the published results at some stage in the past. Although there are still a few small discrepancies, the microdata appears to broadly correspond to the published series, which implies that the microdata can be used in calculating the other sets of variables.

3.2.2 New Indicators

Figure 19 illustrates the results for Total Manufacturing for a selection of questions, comparing the three main weighting procedures. Figure 20 and Figure 21 compares the results for a selection of the new subsector classifications, using the three weighting procedures. Figure 22 presents the Total Manufacturing results for the three largest provinces, using the three main weighting procedures.

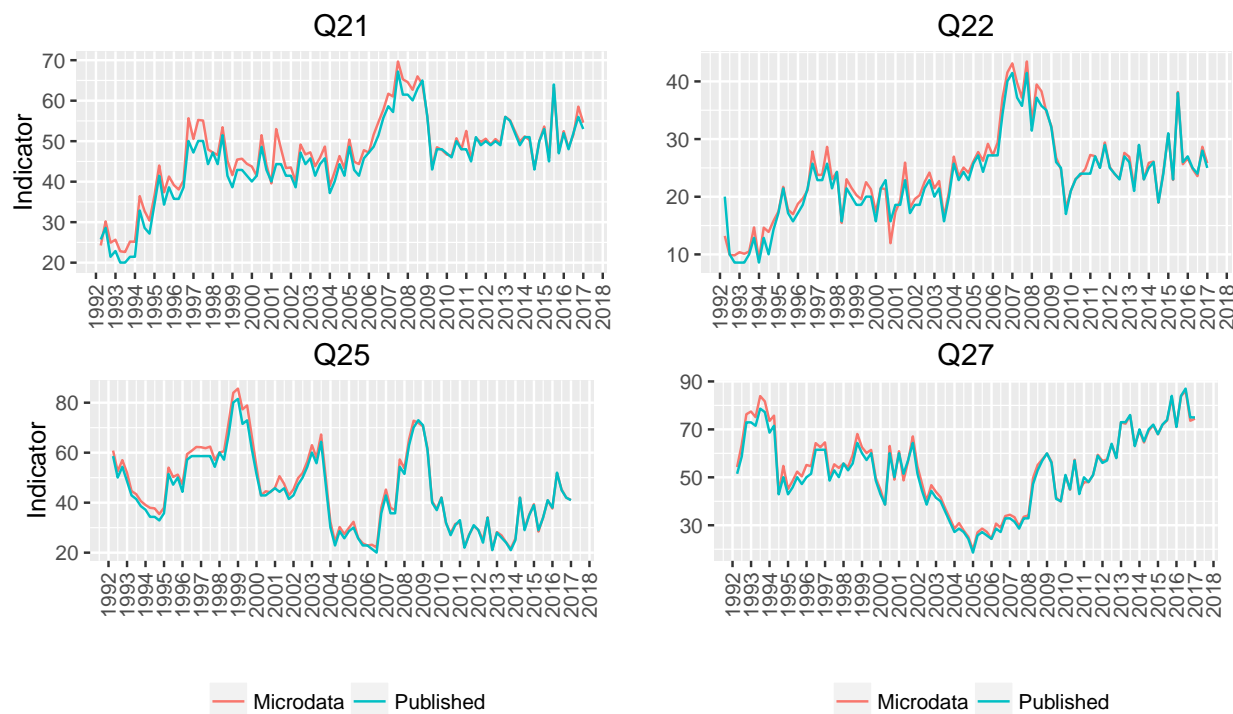


Figure 18: Manufacturing microdata results compared to the published results for constraints

In general the series calculated with the new firm size weights are less volatile than the old series. The two-step procedure produces results that are very similar to those using the old weighting procedure with the new weights. The unweighted series are very similar to these series and generally exhibit the lowest volatility.

3.2.3 Exports

We have also calculated new weights specifically for the survey responses relating to the questions on exports (Q2A, Q2P, Q5A and Q5P). Figure 23 compares the balance series for one of these questions, using the old weights in the microdata, and the two new weighting procedures. The results are quite similar, although the new weights produce less volatile series.

3.3 Reference Series

In this section the series are assessed in terms of their tracking record, i.e. their correlations to their respective reference series. The SARB has two series available for total manufacturing production, based on sales values and volumes. Stats SA also publishes a series for total manufacturing volume. At the subsector level, only the Stats SA series for manufacturing production volumes are available.

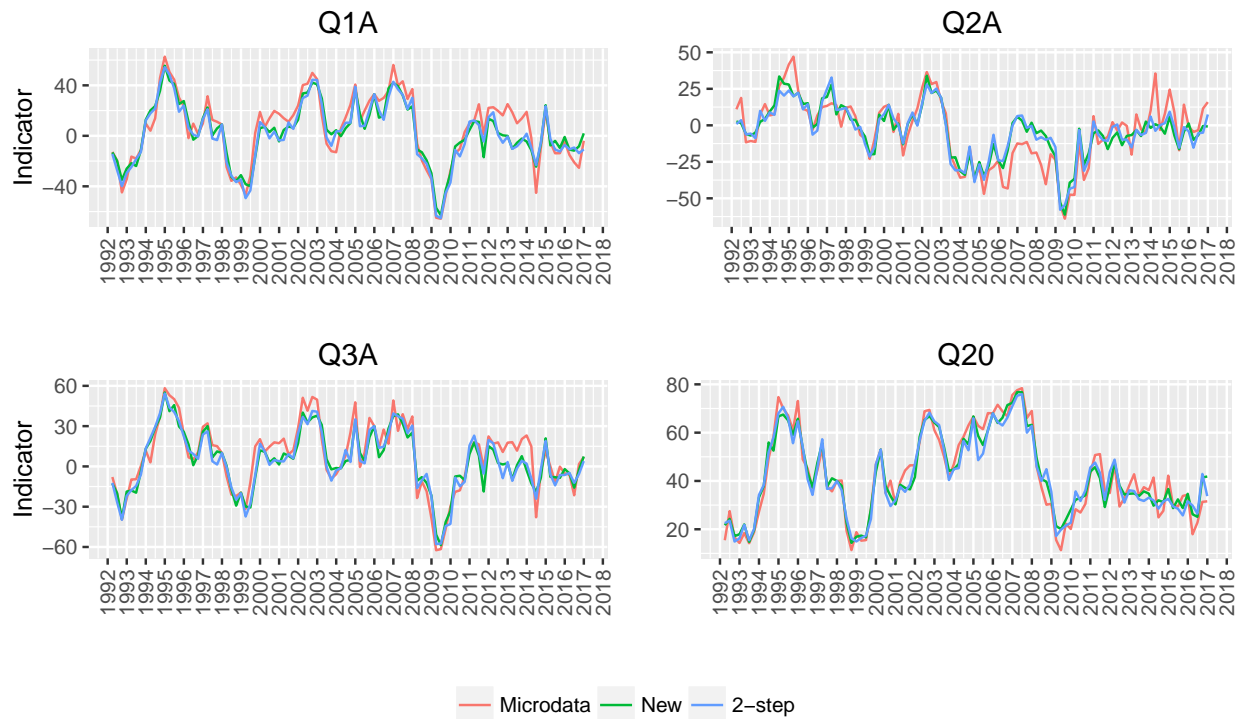


Figure 19: Total Manufacturing results for the alternative weighting procedures

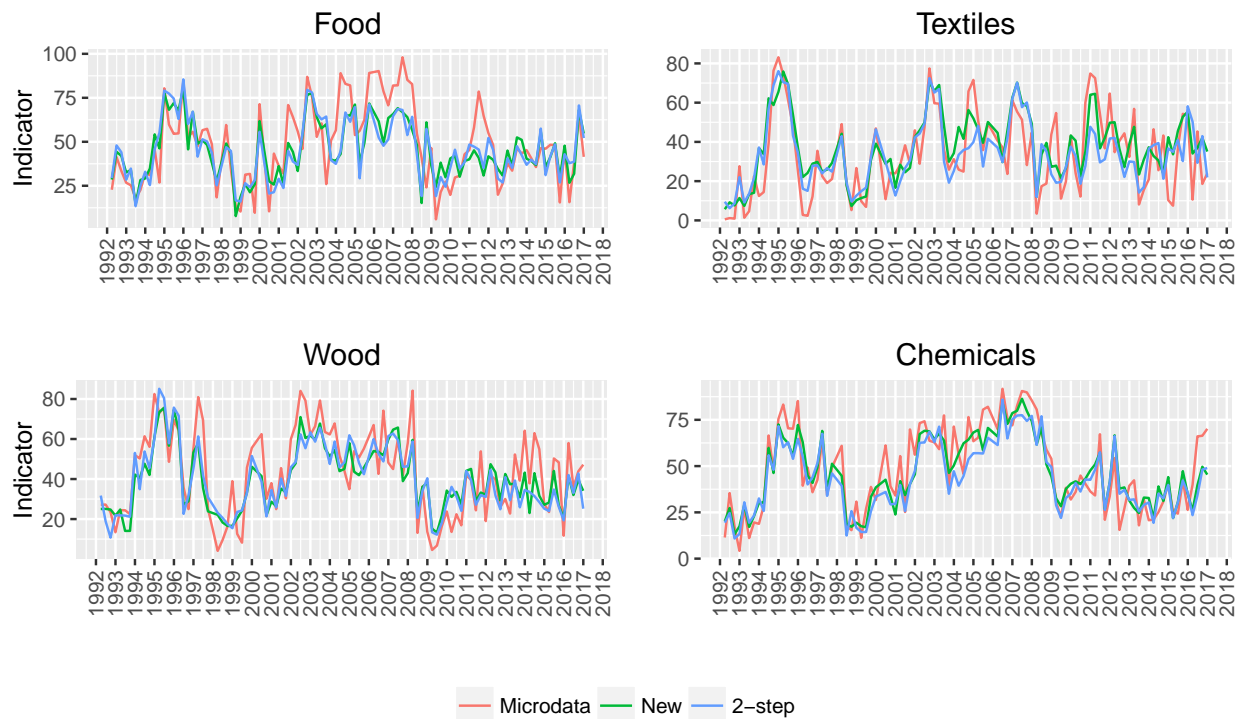


Figure 20: Manufacturing subsector results for the alternative weighting procedures

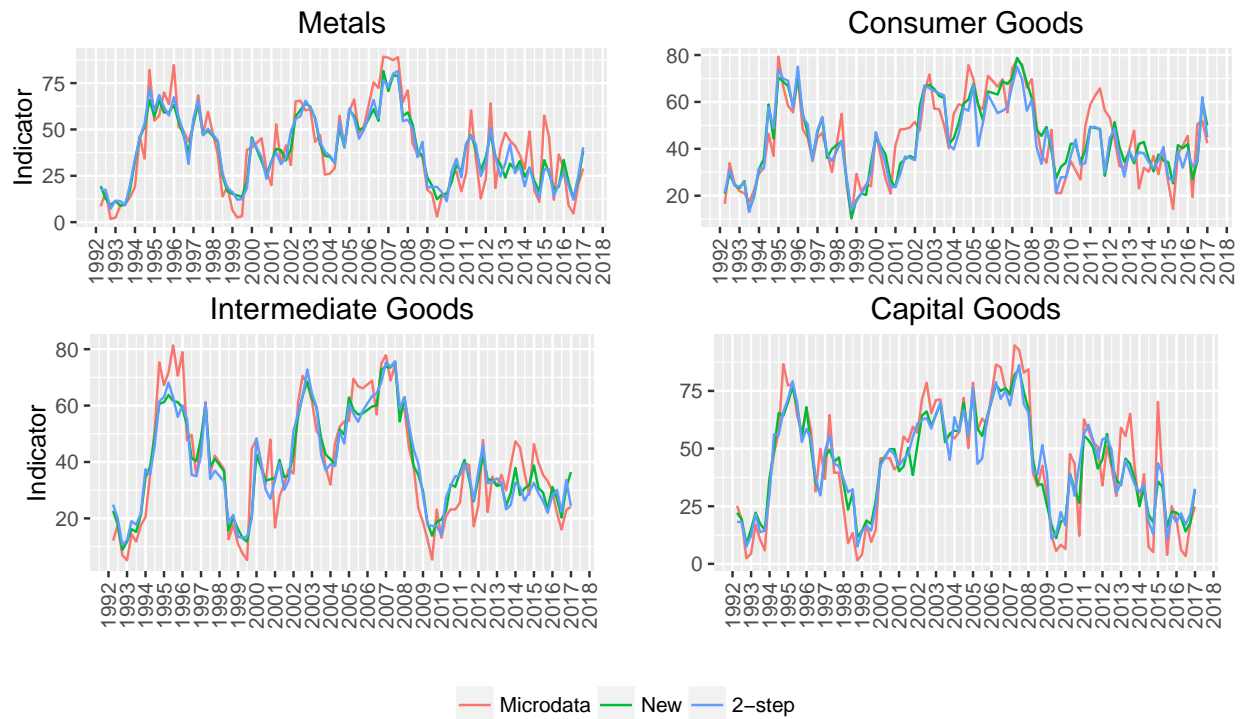


Figure 21: Manufacturing subsector results for the alternative weighting procedures

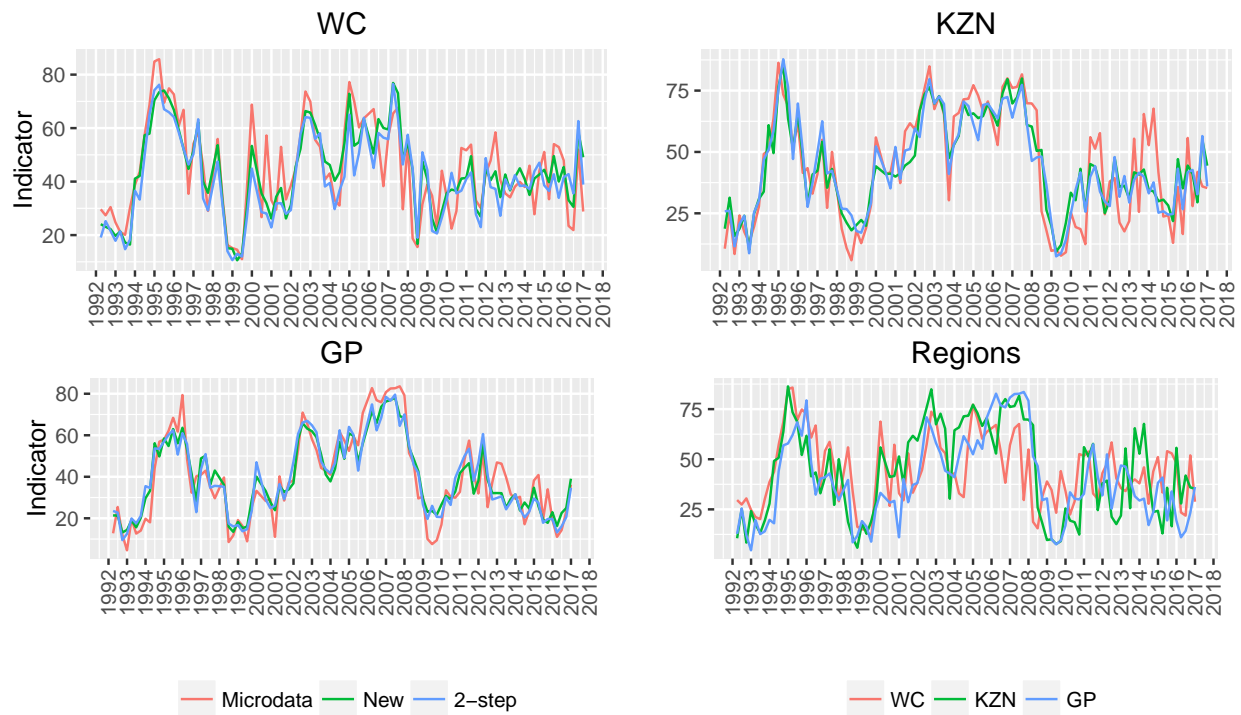


Figure 22: Manufacturing regional results for the alternative weighting procedures

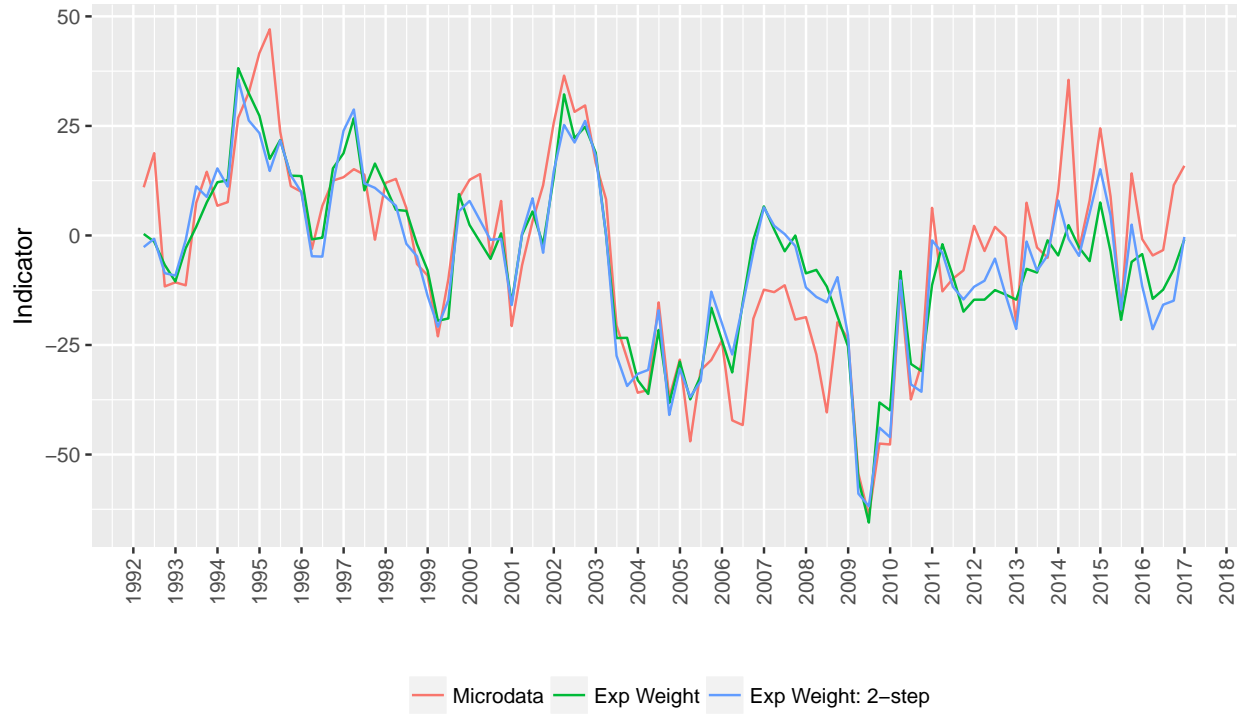


Figure 23: Manufacturing export results for the alternative weighting procedures

3.3.1 Total Manufacturing Activity

Figure 24 illustrates the Total Manufacturing Activity series compared to the SARB and Stats SA references series. Table 10 reports that the contemporaneous correlations between the survey balances and the reference series are relatively high, which is confirmed by the cross-correlations (not shown).

Table 10: Correlations between total manufacturing and reference series

	Q1A	Q1A_u	Q1A_new	Q1A_2s	Q3A	Q3A_u	Q3A_new	Q3A_2s
SARB:Sales	0.62***	0.57***	0.58***	0.59***	0.65***	0.60***	0.61***	0.64***
SARB:Volume	0.72***	0.69***	0.71***	0.70***	0.74***	0.71***	0.72***	0.73***
StatsSA:Volume	0.74***	0.70***	0.72***	0.72***	0.71***	0.70***	0.71***	0.72***

3.3.2 Manufacturing Subsector Activity

Figure 25 and Figure 26 illustrates the Manufacturing subsector activity series compared to the SARB and Stats SA references series. Table 11 reports that the contemporaneous correlations between the survey balances and the reference series are relatively high.

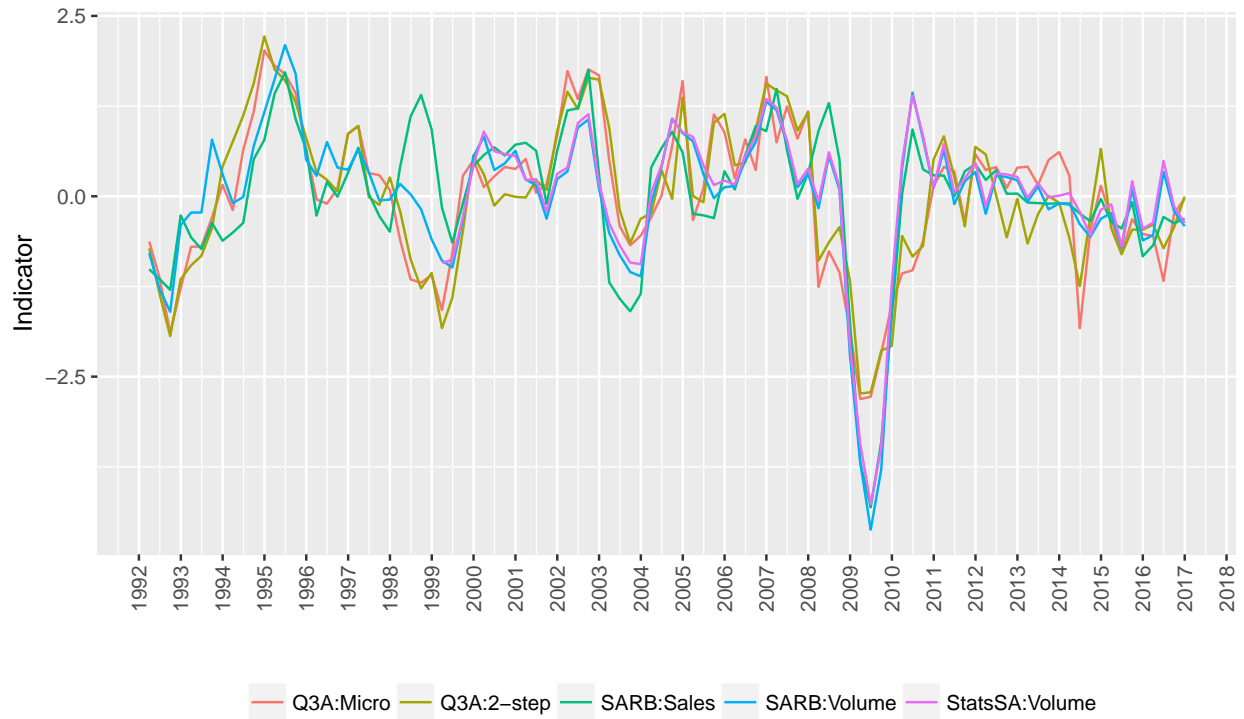


Figure 24: Total Manufacturing compared to the reference series

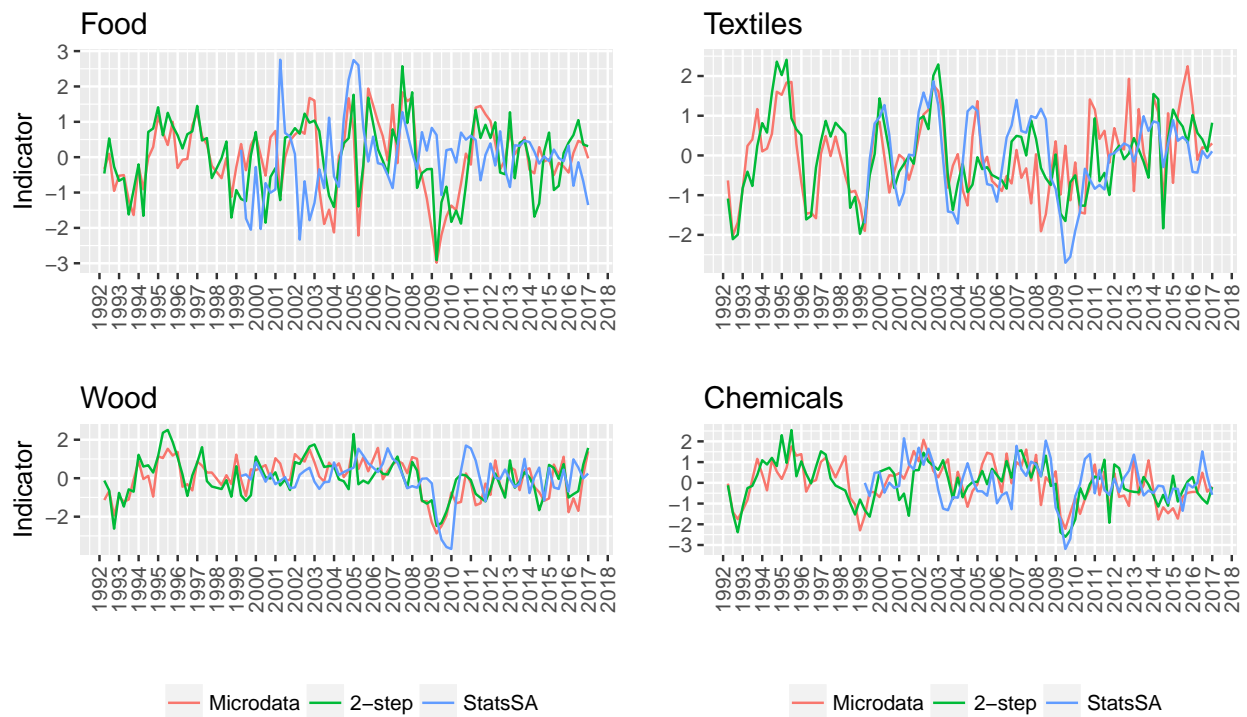


Figure 25: Manufacturing subsectors compared to the reference series

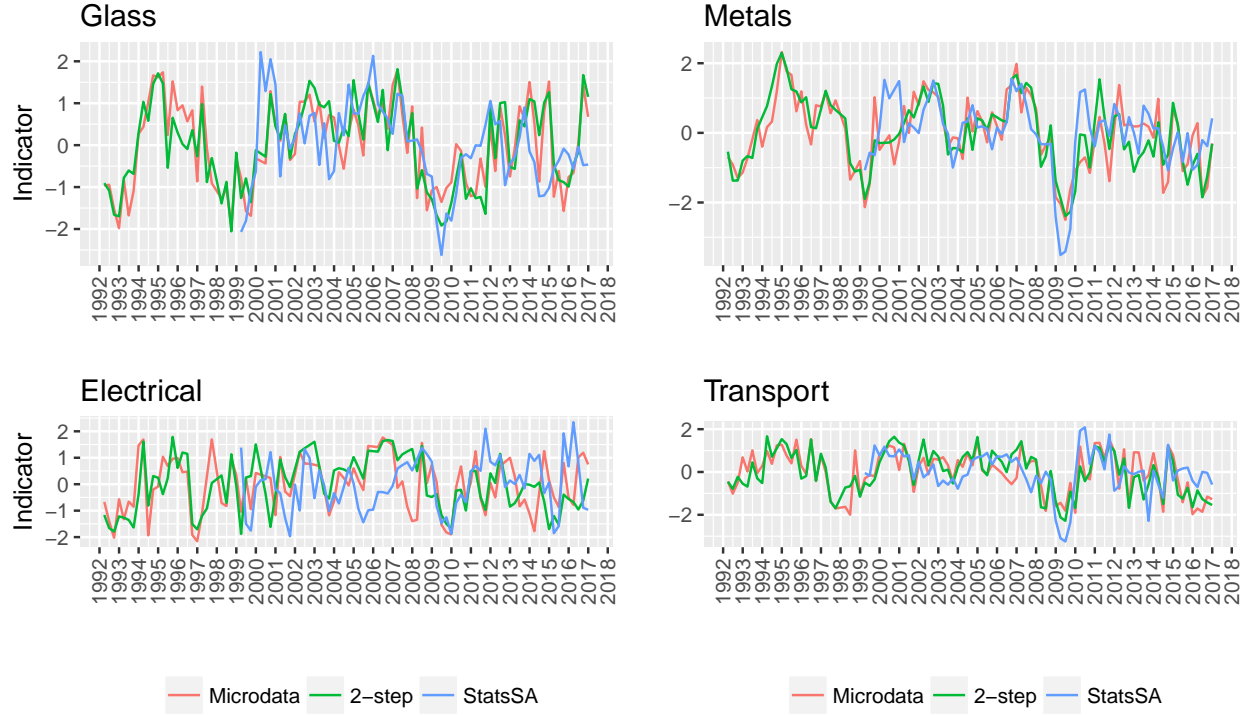


Figure 26: Manufacturing subsectors compared to the reference series

Table 11: Correlations between Manufacturing subsectors and reference series

	Q1A	Q1A_U	Q1A_New	Q1A_2s	Q3A	Q3A_U	Q3A_New	Q3A_2s
Food	-0.08	0.05	0.02	0.01	-0.14	0.05	0.00	-0.02
Textiles	0.29**	0.63***	0.50***	0.63***	0.30**	0.65***	0.51***	0.62***
Wood	0.41***	0.53***	0.47***	0.45***	0.41***	0.49***	0.42***	0.42***
Chemicals	0.31***	0.30**	0.33***	0.36***	0.37***	0.34***	0.38***	0.41***
Glass	0.57***	0.59***	0.61***	0.60***	0.47***	0.56***	0.55***	0.51***
Metals	0.62***	0.68***	0.69***	0.67***	0.57***	0.69***	0.68***	0.66***
Elec_radio	0.01	0.22*	0.03	0.01	0.04	0.24**	0.07	0.07
Transport	0.43***	0.48***	0.48***	0.50***	0.49***	0.47***	0.48***	0.54***
Furniture	0.45***	0.48***	0.48***	0.54***	0.38***	0.41***	0.39***	0.44***

3.3.3 Exports

For the series relating to exports we use the growth rates in the Rand values and Dollar volumes of manufacturing exports as reference series. Figure 27 compares the total export series to the reference series. Figure 28 illustrates the export balances and the reference series for a selection of subsectors. In both cases the reference series are very volatile. There is unrealistic decrease in the total exports reference series in 1994, which is not echoed in the sub-sector series. The subsector export reference series also seem to have unrealistic spikes. The correlations between the balances and the reference series are therefore relatively low, as reported in Table 12, and the correlations with the Dollar volume series are not significant. Still, the new weights seem to improve these correlations.



Figure 27: Total Manufacturing exports compared to the reference series

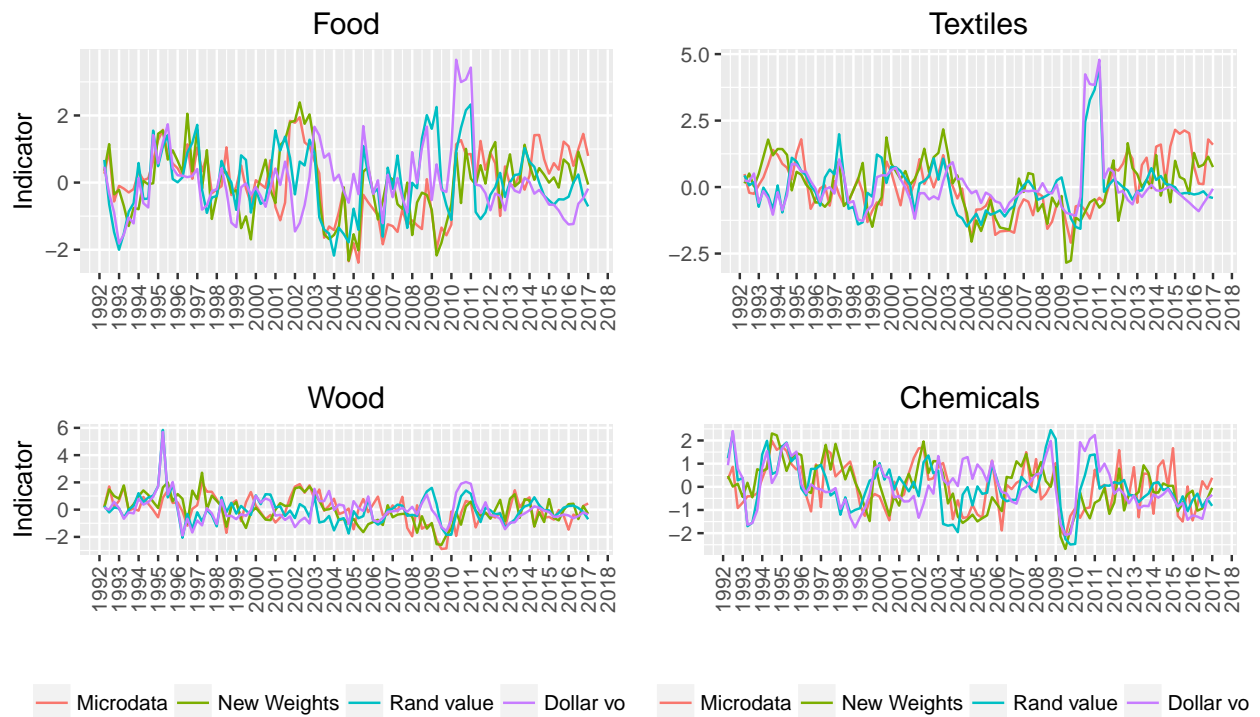


Figure 28: Manufacturing subsector exports compared to the reference series

Table 12: Correlations between the sectoral exports and reference series

	Q2A	Q2A_NExp	Q2A_2Exp	Q5A	Q5A_NExp	Q5A_2Exp
Total_Rval	0.13	0.16	0.22**	0.17*	0.20*	0.25**
Food_Rval	0.30***	0.36***	0.30***	0.31***	0.36***	0.31***
Text_Rval	0.10	0.18*	0.16	0.10	0.16	0.14
Wood_Rval	0.06	0.18*	0.17*	0.08	0.20**	0.21**
Chem_Rval	0.33***	0.37***	0.42***	0.36***	0.39***	0.46***
Glass_Rval	0.11	0.05	0.13	0.08	0.03	0.08
Metals_Rval	0.31***	0.32***	0.37***	0.41***	0.42***	0.44***
Elec_Radio_Rval	0.19*	0.29***	0.23**	0.16	0.26***	0.21**
Motor_Rval	0.05	0.14	0.15	0.02	0.09	0.11
Furn_Rval	0.20**	0.18*	0.19*	0.18*	0.12	0.12

3.3.4 Volatility

We can also compare the series in terms of volatility. Table 13 reports the volatilities of the series over time. In general the new weights produce series that are less volatile than the old series. The unweighted series are often the least volatile, implying that the weights impart some volatility to the series.

Table 13: Volatility of the Manufacturing series

	Q1A	Q1A_U	Q1A_New	Q1A_2s	Q3A	Q3A_U	Q3A_New	Q3A_2s
Total	27.18	22.99	23.20	24.45	24.98	21.58	21.71	22.66
Food	35.81	22.77	24.00	26.79	35.76	20.65	23.21	25.05
Textiles	36.12	26.12	26.96	27.43	36.62	26.24	27.25	28.45
Wood	36.86	25.63	25.96	27.93	33.28	24.45	24.67	26.01
Chemicals	32.16	23.39	24.16	24.75	30.83	22.37	22.59	23.04
Glass	46.02	31.72	36.53	34.94	43.94	28.72	32.03	31.08
Metals	37.79	28.06	29.60	30.59	34.96	27.29	28.18	29.36
Elec_radio	50.29	33.03	48.84	48.98	49.30	33.03	49.23	50.63
Transport	55.88	32.70	38.78	43.61	52.29	32.52	38.30	41.92
Furniture	44.71	31.50	35.29	35.01	43.07	30.42	34.10	33.45

3.4 Summary

After some corrections the microdata seems to provide series that correspond very closely to the published series. We have calculated new series, using four weighting procedures, and compared them to the reference series. We chose the Stats SA series as reference series, as it is widely available for many of our subsectors and has a relatively high correlation with the balances. The final series still have to be written to an Excel file in a format that is convenient for processing. The updates can and should be fully automated.

4 The Trade Surveys

In this section we examine the Trade sector surveys: Retail, Wholesale and Motor Vehicles. The Retail, Wholesale and Motor Vehicle microdata is available from 1992Q2-2017Q2. Similar to the other surveys, to match the questions and responses in the different versions of the surveys, the surveys in the old records were compared to the current incarnations. We identified and marked the questions in the old (pre-2001) surveys and matched them to the new survey questions. The old questionnaires have been uploaded to the Dropbox folder.

Where published series were available, we used the published values for those missing values. Where they were unavailable, we used simple linear interpolation to create complete continuous series. There is very little difference between the substitution of the published values and simple linear interpolation. In all cases we corrected the clear errors in the microdata. Some of the original sector codes were not included in the new sector classifications and these were reclassified to correspond to the new sector code classifications.

4.1 The Retail Survey

This section presents the results for the Retail surveys. Table 14 reports some characteristics of the Retail survey microdata. The Retail microdata are missing for the following quarters: 1997Q4, 2000Q1 and 2005Q4. Table 15 reports the new subsector codes that are used to calculate balances. Balances are calculated for *total retail*, *durable*, *semi-durable* and *non-durable goods*. In addition, we calculate separate series for *hardware* and *other durable goods*. Figure 29 illustrates the number of respondents by subsector, according to these sector classifications.

Table 14: Sample characteristics

	Sample	Total Obs	Obs/Quarter	Missing Quarters
Retail Survey	1992Q2-2017Q2	20008	204.16	1992Q4,1993Q3,2005Q4

4.1.1 The Retail Indicators

Similar to the Manufacturing survey we have created three versions of the balance statistics for all of the questions. The baseline balance statistics are calculated in the same way as the normal BER procedure. The weights are calculated as: $w_{it} = f_{it}s_{jt}$, where f_{it} the firm size weight (i.e. the inner weight reflecting turnover or number of employees) for firm i at time t ; and s_{jt} is the subsector weight (i.e. the outer weight reflecting the share of total value added) for subsector j at time t . These series are then compared to the BER's published series to ensure that the microdata is correct.

Table 15: Retail Sector Codes

Sector	Subsector	Code
Semi-Durables		3110, 3120, 3160, 3140, 3130, 3170, 3150
Non-Durables		3230, 3210, 3240, 3220
Durables	Hardware	3330
	Other	3370, 3310, 3340, 3350, 3320, 3380, 3360
Other		3410, 3420

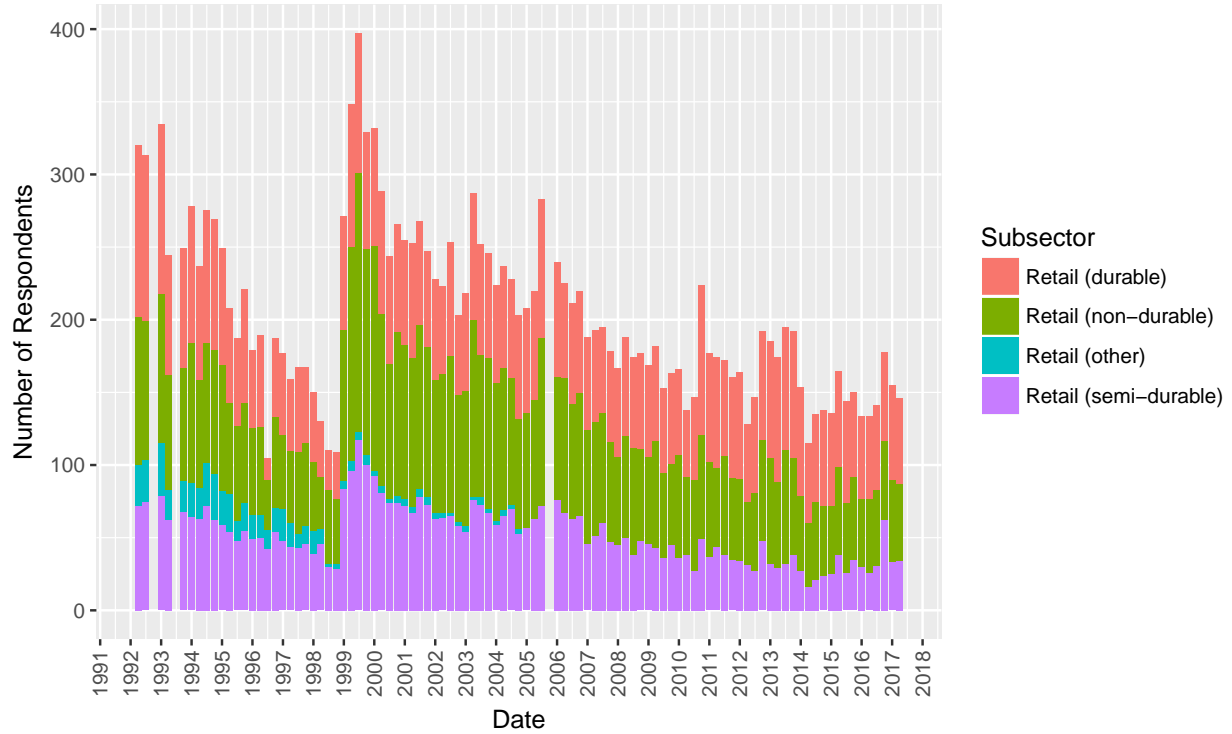


Figure 29: The number of respondents by subsector (1992Q2-2017Q2)

The second set of indicators is based on a new 4-part firm classification, which are less exponential than the previous weights. The final set of indicators is based on the two-step weighting procedure, where the weights are calculated as: $w_{it} = f_{it}s_{jt}/F_{jt}$, where f_{it} the firm size weight (i.e. the new 4-part classification) for firm i at time t ; s_{jt} is the subsector weight for subsector j at time t ; and $F_{jt} = \sum_{i=1}^N f_{it}$ is the total firm weight for subsector j at time t .

4.1.1.1 Published Series

In this section the published results are compared to the balances calculated with the microdata. However, the comparison is complicated by the revisions to the published results in 1997Q1 and 2008Q3, when the sector weights were adjusted. To chain the old and new series together, an adjustment factor was added to all the quarters before the revisions. The adjustment factor was calculated as the difference between the results using the existing and revised weights in the quarter when the revision was done. This adjustment was then added to the historical series.

To make a comparison, therefore, the adjustment factors were subtracted from the published series for comparison to the microdata. The adjustment factors are available in the Excel files on Dropbox. Thus, the microdata series should correspond to the published series before the revisions and adjustment factors.

Figure 30 illustrates the results for total Retail, comparing the published series (with and without adjustment) to the balances based on the microdata. Figure 31 illustrates the comparison for a selection of questions. The series are almost identical, except for 1997Q2-1998Q2 and 2010Q4. Figure 32 illustrates the comparisons for the three main regions. The series are very similar, especially

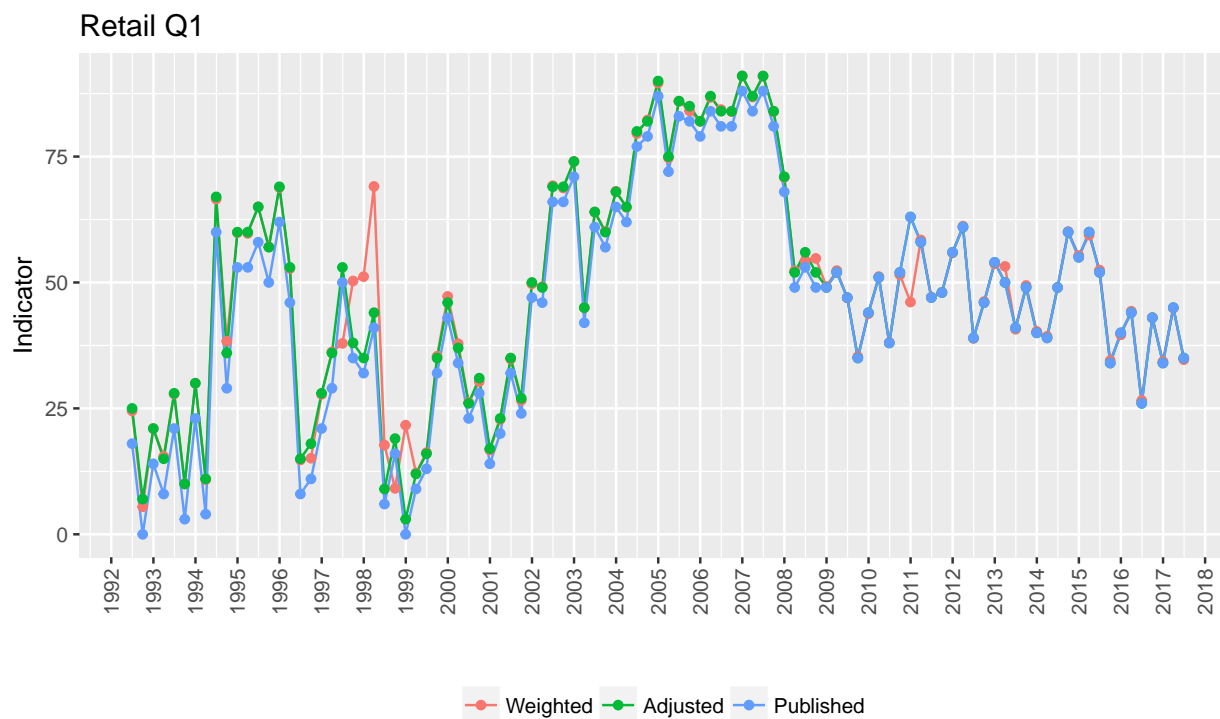


Figure 30: Total Retail compared to the published series

after 2001. The microdata therefore seems to correspond to the published series (before adjustment factors), and the microdata therefore seems to be correct.

4.1.1.2 New Indicators

This section compares the microdata results using the three different weighting procedures. Figure 33 illustrates the comparison for Total Retail for a selection of questions, comparing the three main weighting procedures. Figure 34 compares the results for a selection of the new Retail subsector classifications, using the three weighting procedures. The new weights produce series that are less volatile than the old series.

4.1.2 Reference Series

This section compares the series to their respective reference series. Stats SA and the SARB calculate series for total retail sales, but the old retail series ceased in 2003 and only resumed again in 2008 (constant prices). George used additional data to fill in this gap, to create a continuous complete reference series, at least for total retail sales. Unfortunately, we do not have a complete reference series for the subsectors, which complicates the comparison and the evaluation of the tracking record. The “other” component is not allocated to the durable, semi-durable and non-durable sectors, as it does not have a large impact on the growth rates of the components.

Figure 35 compares the Total Retail activity series (Q2A and Q3A) to the SARB and Stats SA reference series. Table 16 reports the contemporaneous correlations between the series. The

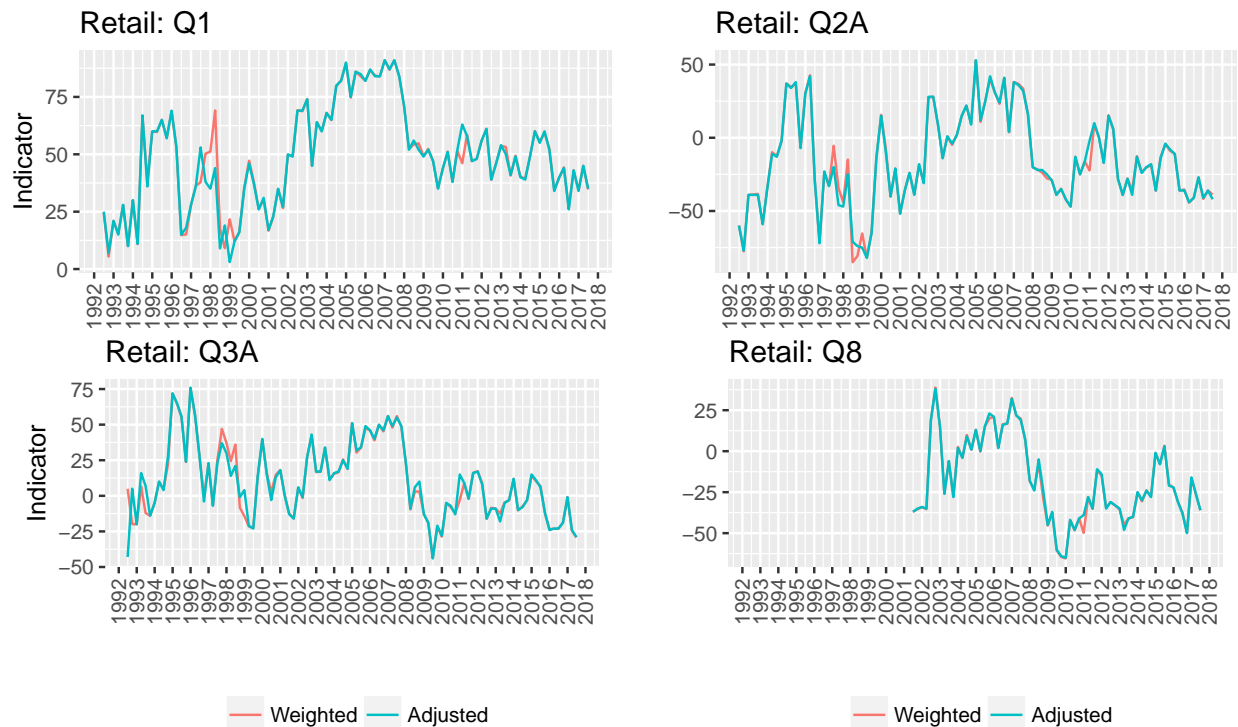


Figure 31: Retail subsectors compared to the published series

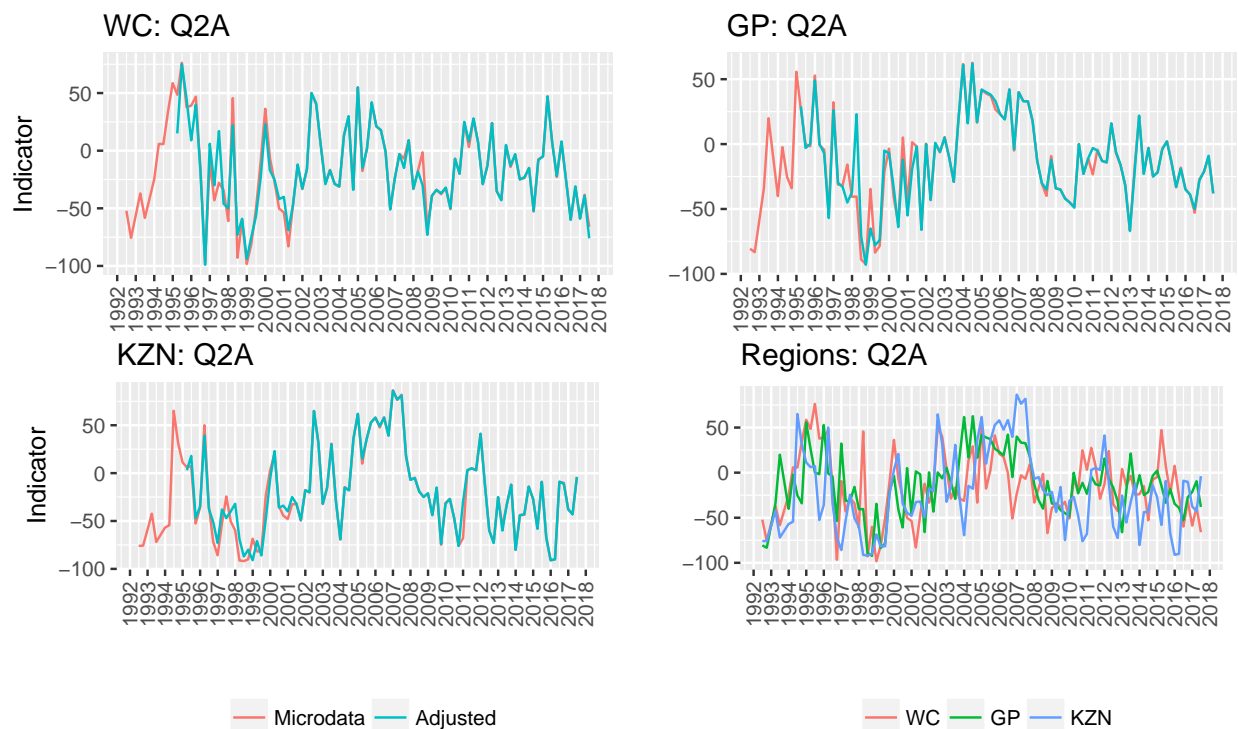


Figure 32: Regional Retail compared to the published series

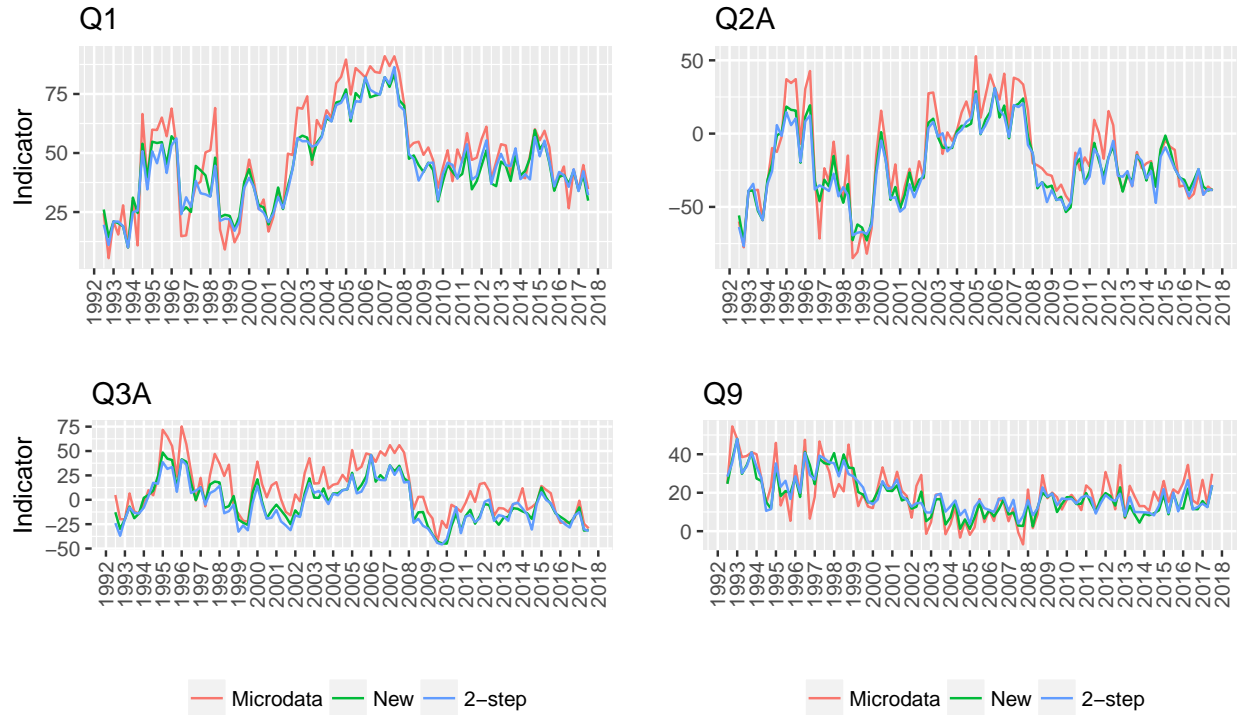


Figure 33: Total Retail series using different weights

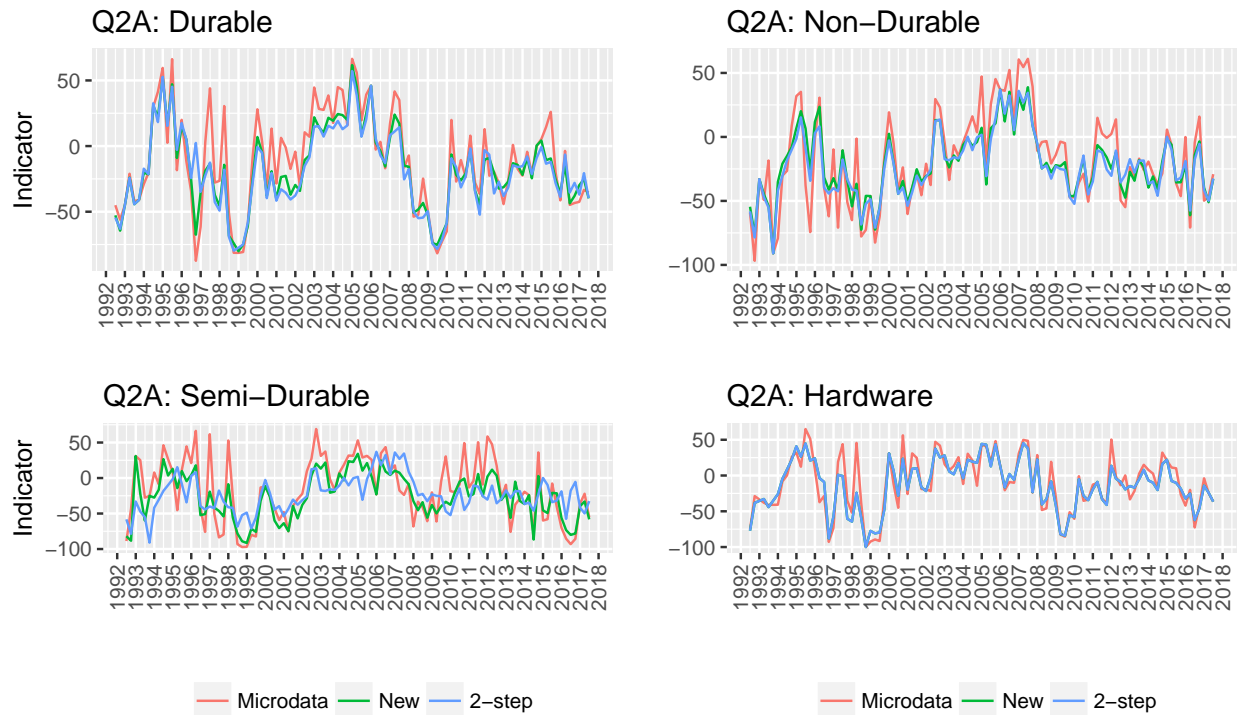


Figure 34: Retail subsectors using different weights

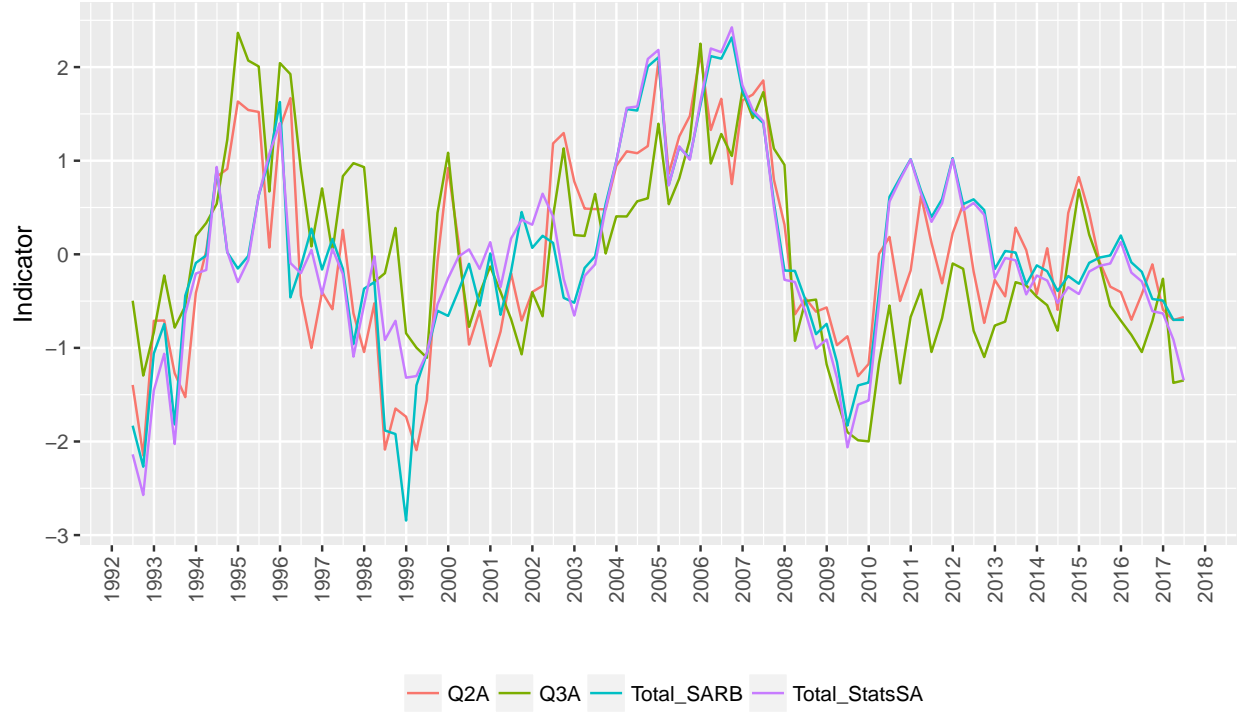


Figure 35: Total Retail compared to the reference series

correlations are generally relatively high and significant. This significant positive correlation is confirmed by the cross-correlations illustrated in Figure 36. The new weights also seem to improve the tracking record of the balance statistics, compared to the old weights from the published series.

Table 16: Correlations of retail series

	Q2A	Q2A_u	Q2A_new	Q2A_2s	Q3A	Q3A_u	Q3A_new	Q3A_2s
Total_SARB	0.71***	0.76***	0.74***	0.77***	0.51***	0.57***	0.52***	0.59***
Total_StatsSA	0.69***	0.74***	0.73***	0.75***	0.56***	0.57***	0.55***	0.60***

Figure 37 compares the three retail subsectors to their reference series. Notice that there are missing values in the reference series between 2003 and 2008, which confounds the comparison to some extent. Nevertheless, Table 17 reports that the correlations for Durable and Semi-Durable goods are still relatively high and significant. The new weights again seem to improve this tracking record, even to the point where one of the series for Non-Durable goods has a somewhat significant positive correlation to its erratic reference series.

Table 17: Correlations of Retail subsector series

	Q2A	Q2A_n	Q2A_2s	Q3A	Q3A_n	Q3A_2s
Durable Goods	0.57***	0.60***	0.61***	0.61***	0.61***	0.61***
Semi-Durable Goods	0.43***	0.51***	0.49***	0.46***	0.49***	0.47***
Non-Durable Goods	0.13	0.25**	0.28**	-0.14	-0.09	-0.06

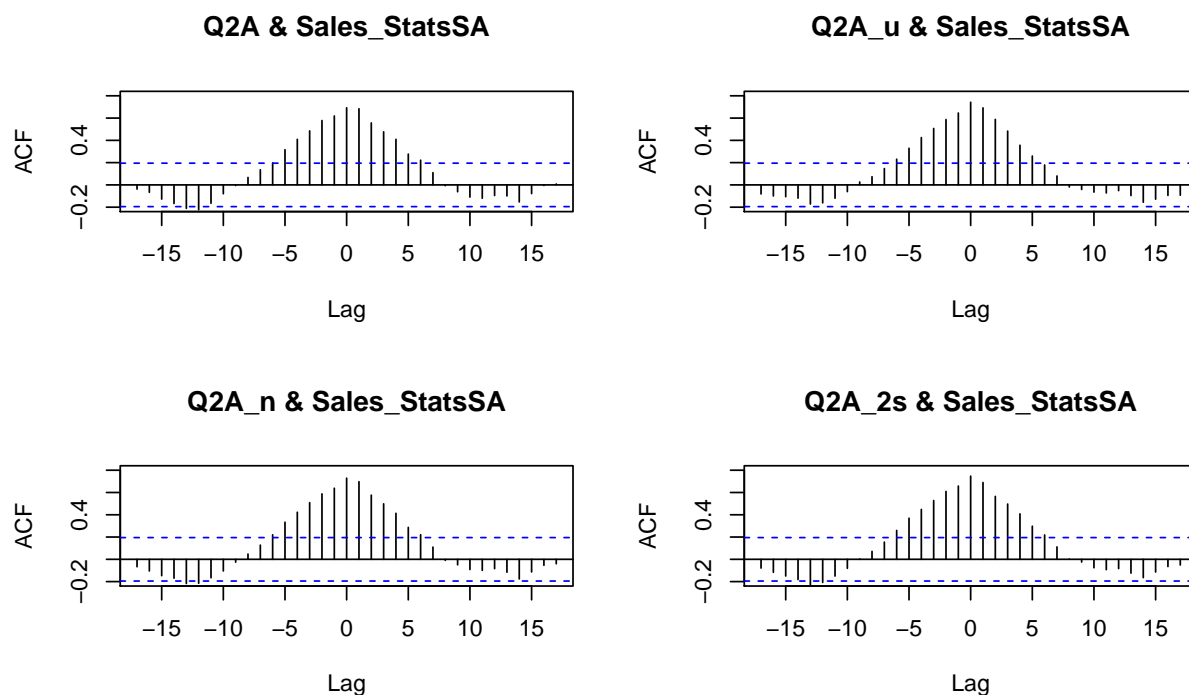


Figure 36: Retail cross-correlations to the reference series

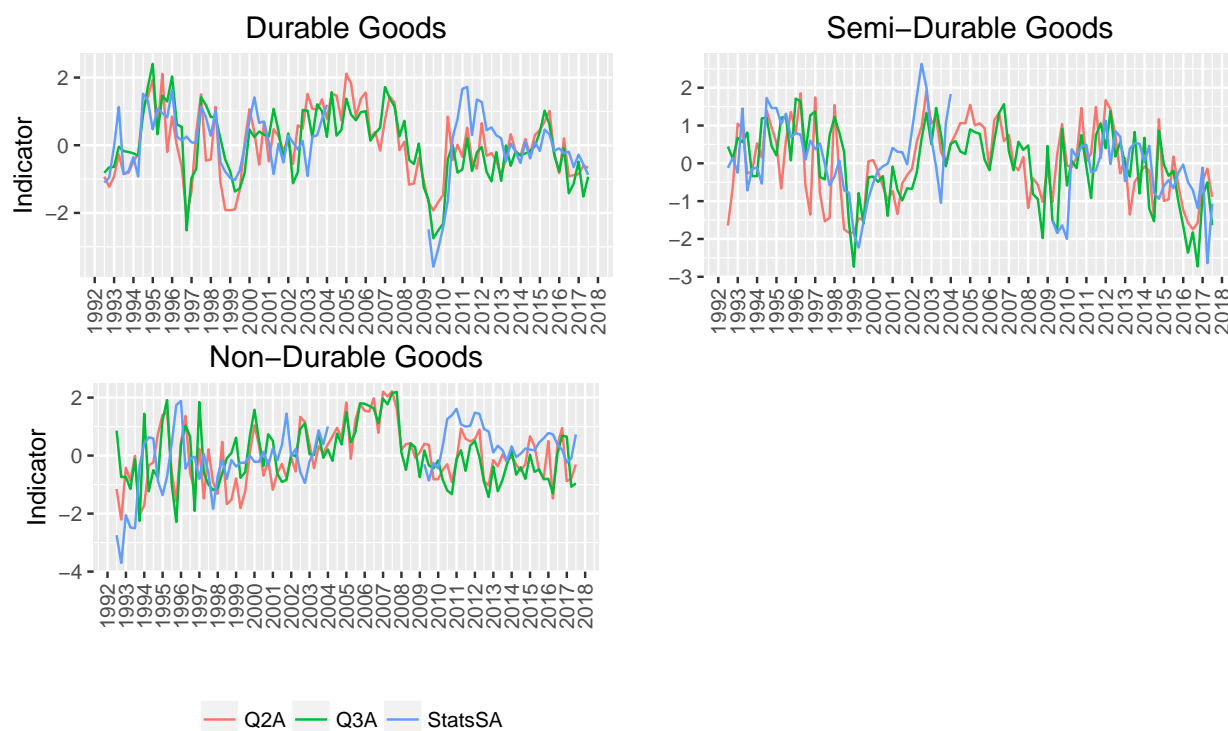


Figure 37: Retail subsectors compared to the reference series

4.1.3 Optimum Weights

In this section we investigate whether there are more optimum weights that can be used to combine the retail subsector series, in order to improve the tracking record of the total retail series. We briefly investigate three possibilities. The first is to combine the subsector series using simple averages (i.e. equal weights). The second is to use principal components analysis to combine the series. PCA is used to reduce the dimensionality of a dataset consisting of a large number of variables, while retaining as much of the variation as possible (Jolliffe, 2002). The transformation is defined in such a way that the first principal component accounts for as much of the variability in the data as possible. We therefore use the first principal component as a combined measure of the total retail series. The third possibility is to use regression analysis to calculate a predicted series for total retail.

Table 18 reports the results from the regression analysis using the complete subsector series. The results show that most contemporaneous sector series are not significant, due to their high standard errors. Only subsectors 3160 exhibits a significant positive coefficient. Subsectors 3310, 3160 and 3230 receive the largest weights, while one sector (3130) has a slightly negative weight.

Table 18: Regression output

	<i>Dependent variable:</i>
	Retail_ref
'3310'	0.015 (0.012)
'3160'	0.023** (0.012)
'3230'	0.027 (0.017)
'3120'	0.007 (0.008)
'3330'	0.011 (0.012)
'3130'	-0.0002 (0.006)
'3140'	0.005 (0.009)
'3220'	0.003 (0.017)
'3370'	0.003 (0.010)
'3210'	0.008 (0.007)
'3110'	0.011 (0.012)
Constant	6.257*** (0.504)
Observations	99
R ²	0.576
Adjusted R ²	0.523
Residual Std. Error	2.815 (df = 87)
F Statistic	10.759*** (df = 11; 87)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Figure 38 illustrates these series along with the total retail sales reference series from Stats SA. The three methods seem to produce very similar results to the series derived with the new survey results. Table 19 reports that they have high correlations with one another. They also improve the tracking record with the reference series to some degree, as expected. Nevertheless, our new weights and the 2-step weighting procedure seem to be an adequate way of summarising the survey responses in this case.

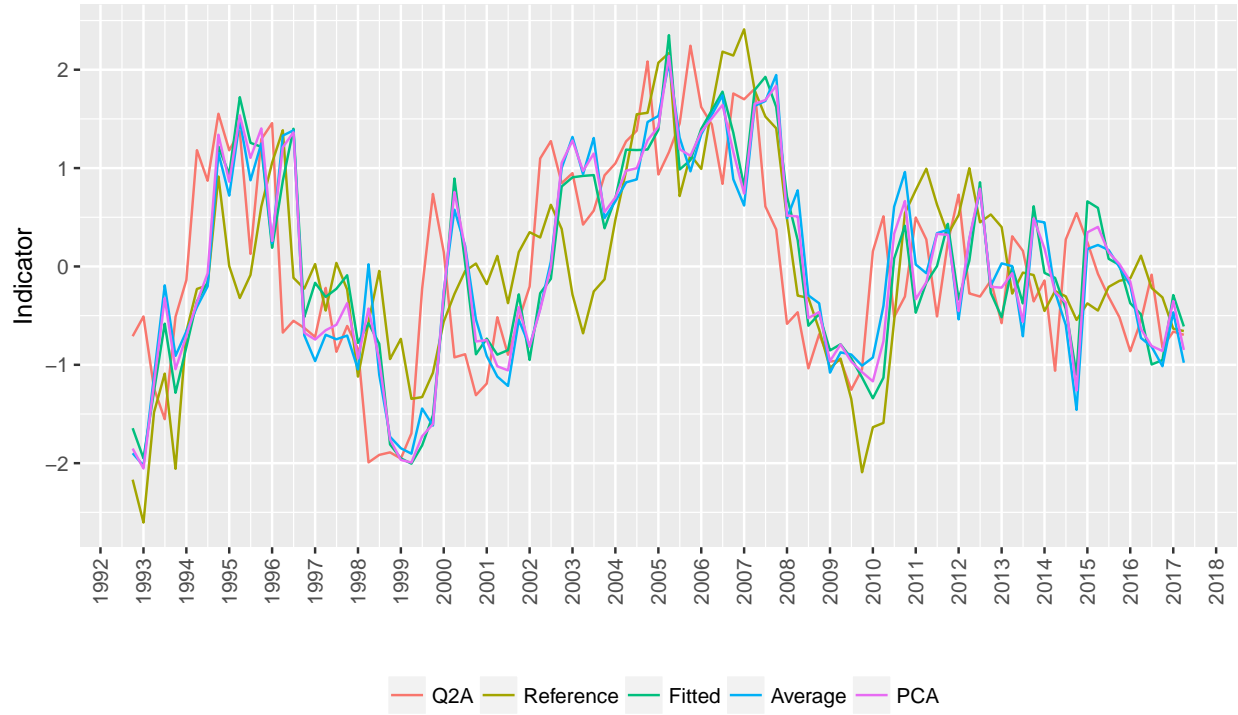


Figure 38: Retail subsectors compared to the reference series

Table 19: Correlations between the Total Retail Series

	Reference	Q2A	Fitted	Average
Q2A	0.61***			
Fitted	0.76***	0.65***		
Average	0.73***	0.62***	0.96***	
PCA	0.75***	0.65***	0.98***	0.99***

4.1.4 Volatility

Finally, we can compare the series in terms of volatility. Table 20 reports the volatilities of the series over time. The new weights and weighting procedures again produce balance series that are substantially less volatile than the original series.

Table 20: Volatility of Retail series

	Q2A	Q2A_New	Q2A_2s	Q3A	Q3A_New	Q3A_2s
Total Retail	31.16	24.51	24.43	25.78	21.40	21.01
Durable Goods	36.66	30.79	30.07	33.91	27.55	27.90
Semi-Durable Goods	44.38	32.17	31.26	38.55	29.68	29.30
Non-Durable Goods	35.68	25.20	24.73	29.65	21.86	21.18

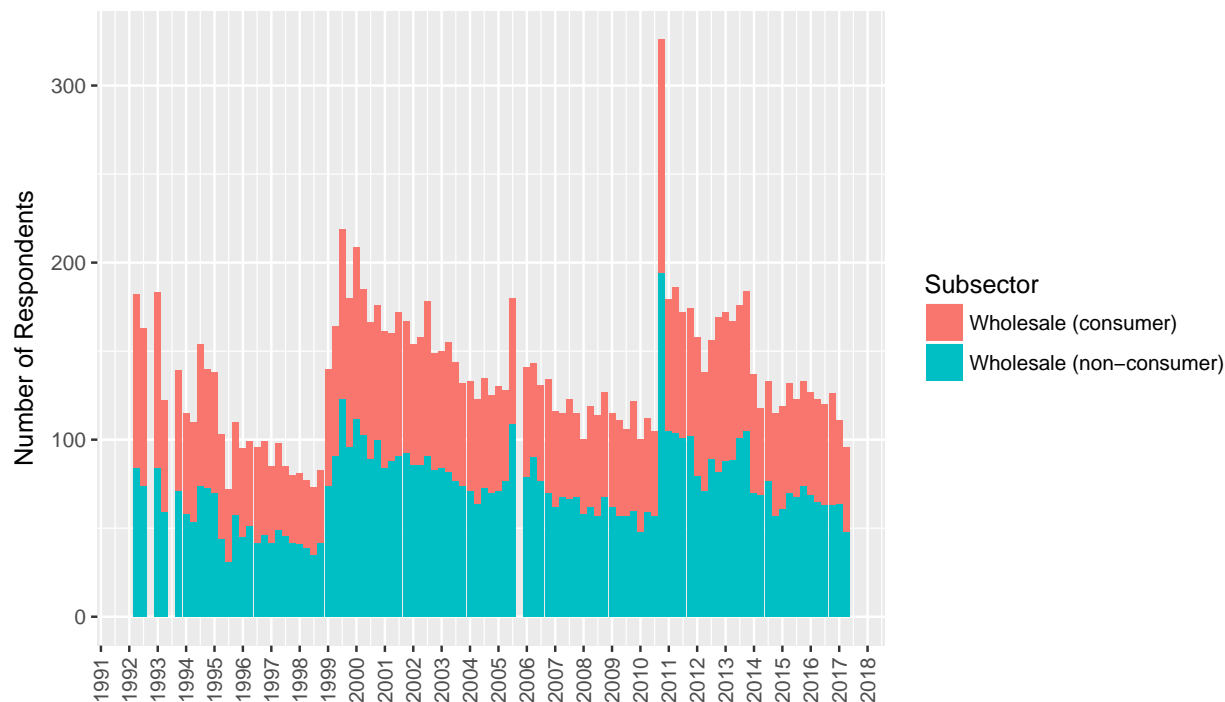


Figure 39: Number of respondents in the wholesale sector (1992Q2-2017Q2)

4.2 The Wholesale Survey

Table 21: Sample characteristics

	Sample	Total Obs	Obs/Quarter	Missing Quarters
Wholesale Survey	1992Q2-2017Q2	13378	136.51	1992Q4,1993Q3,2005Q4

4.2.1 The Wholesale Indicators

Similar to the Manufacturing and Retail surveys we have created three versions of the balance statistics for all of the questions. The baseline balance statistics are calculated in the same way as the normal BER procedure. The second set of indicators is based on a new 4-part firm classification, although in the case of wholesale none of the firms are classified as large (category 4). The final set of indicators is based on the two-step weighting procedure, using the new 4-part firm size weights.

4.2.1.1 Published Series

Table 22: Wholesale Sector Codes

Sector	Code
Non-Consumer Goods	2120, 2110, 2130, 2140
Consumer Goods	2250, 2220, 2230, 2240, 2210, 2270, 2260

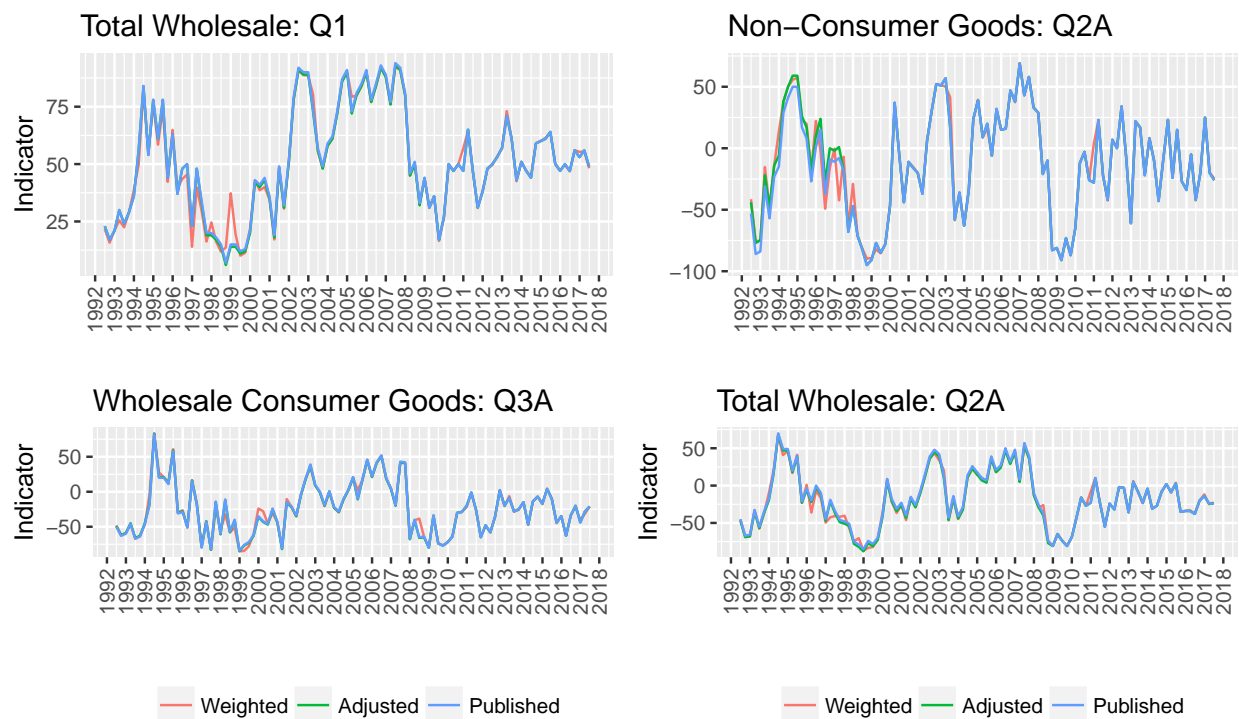


Figure 40: Wholesale subsectors compared to the published series

In this section the published results are compared to the balances calculated with the microdata. Similar to the Retail surveys, however, the comparison is complicated by the revisions to the published results in 1997Q1 and 2008Q3. To make a comparison, the adjustment factors were again subtracted from the published series for comparison to the microdata. The adjustment factors are available in the Excel files on Dropbox.

Figure 40 illustrates the results for a selection of questions for total wholesale, consumer and non-consumer goods. In this case, the adjustments were relatively minor and the series seem to correspond well to the published data, except for a few minor deviations. Figure 41 illustrates the comparisons for the three main regions. The series are very similar, especially after 2001. The microdata therefore seems to correspond to the published series, and the microdata therefore seems to be correct.

4.2.1.2 New Indicators

This section compares the microdata results using the three different weighting procedures. Figure 42 illustrates the comparison for Total Wholesale for a selection of questions, comparing the three main weighting procedures. Figure 43 compares the results for a selection of the new Wholesale subsector classifications, using the three weighting procedures. The new weights produce series that are less volatile than the old series.

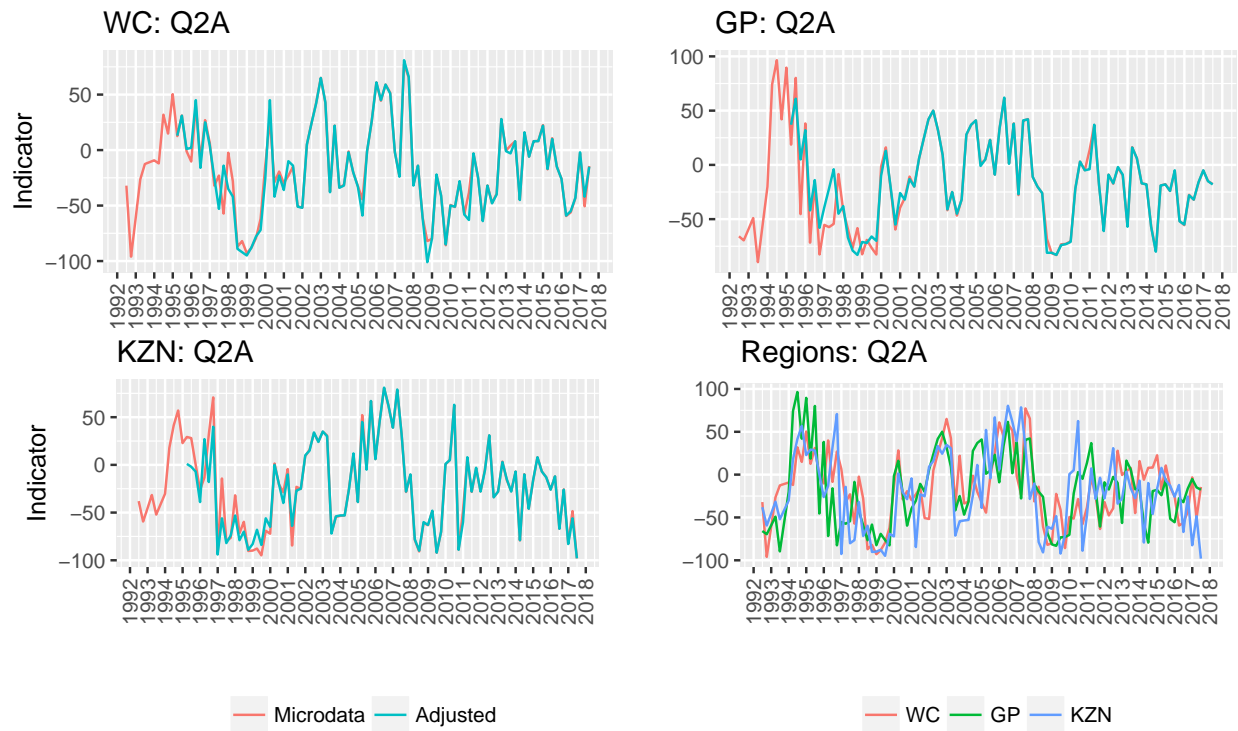


Figure 41: Wholesale regions compared to the published series

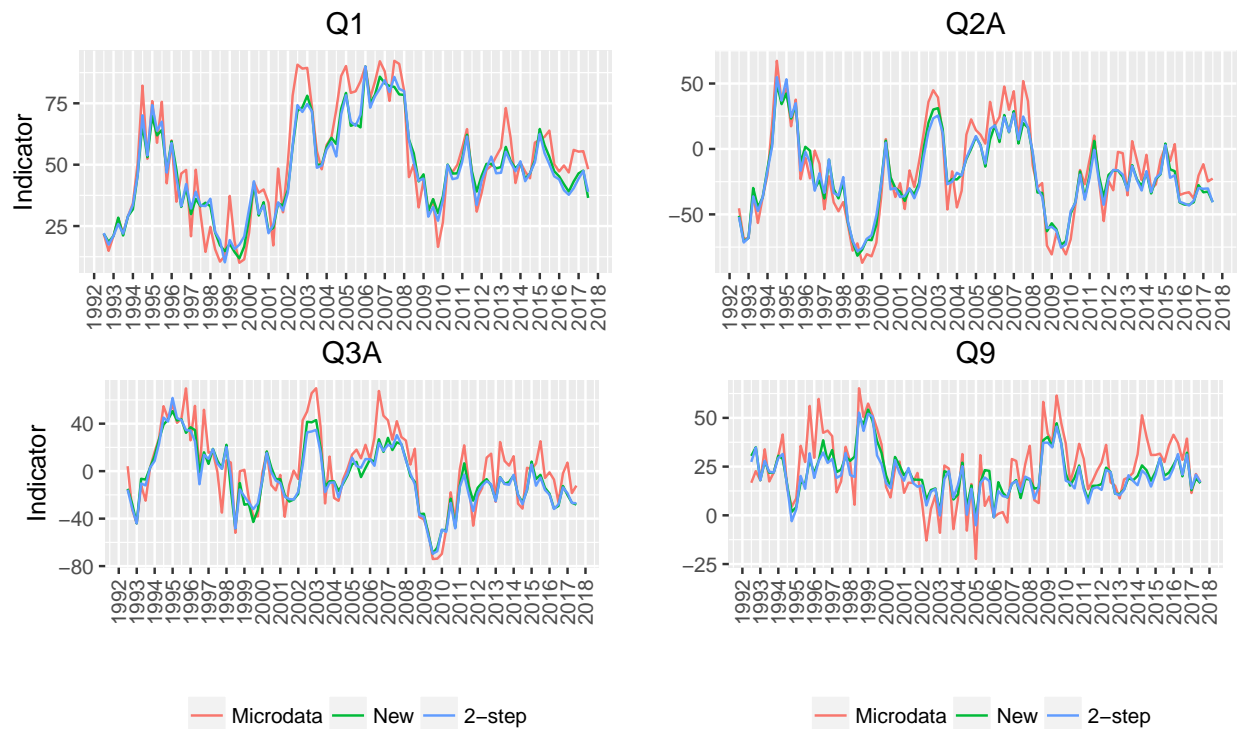


Figure 42: Total Wholesale series using different weights

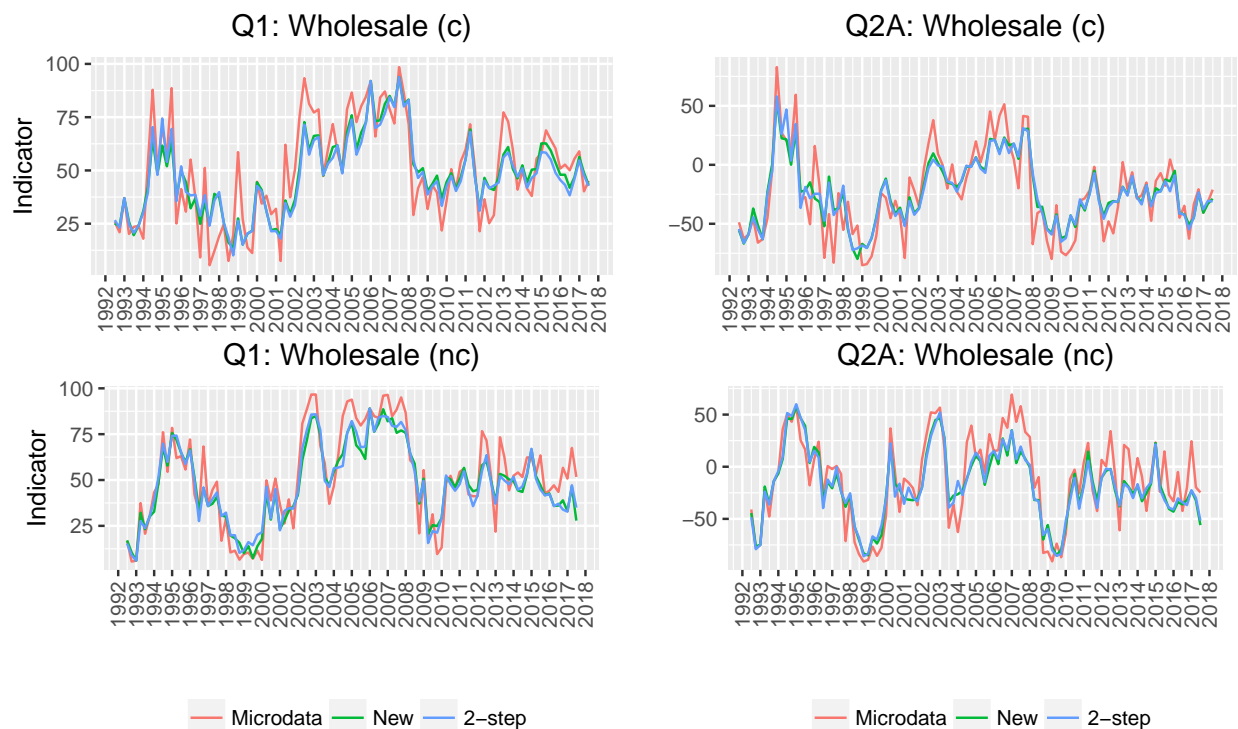


Figure 43: Wholesale subsectors using different weights

4.2.2 Reference Series

We still need to choose reference series for the wholesale sector. The Wholesale sector data is only available in current prices since 2005. Stats SA’s sector classification has a large “other” component, which may have a large impact on the growth rates. Once we have chosen reference series we can evaluate the wholesale series in terms of tracking record and volatility.

4.3 The Motor Vehicle Survey

This section presents the results for the Motor Vehicle surveys. Table 23 reports some characteristics of the Motor survey microdata. The Motor microdata is available from 2005Q2-2017Q2 and 2005Q4 is missing. There is only one sector in these survey, i.e. 3510, so Figure 44 illustrates the number of respondents by region.

Table 23: Sample characteristics				
	Sample	Total Obs	Obs/Quarter	Missing Quarters
Motor Vehicle Survey	1992Q2-2017Q2	6389	65.19	1992Q4,1993Q3,2005Q4

4.3.1 The Motor Vehicle Indicators

Similar to the previous surveys, we calculate balance statistics in the same way as the normal BER procedure, in order to compare the results to the published series. The second set of indicators is

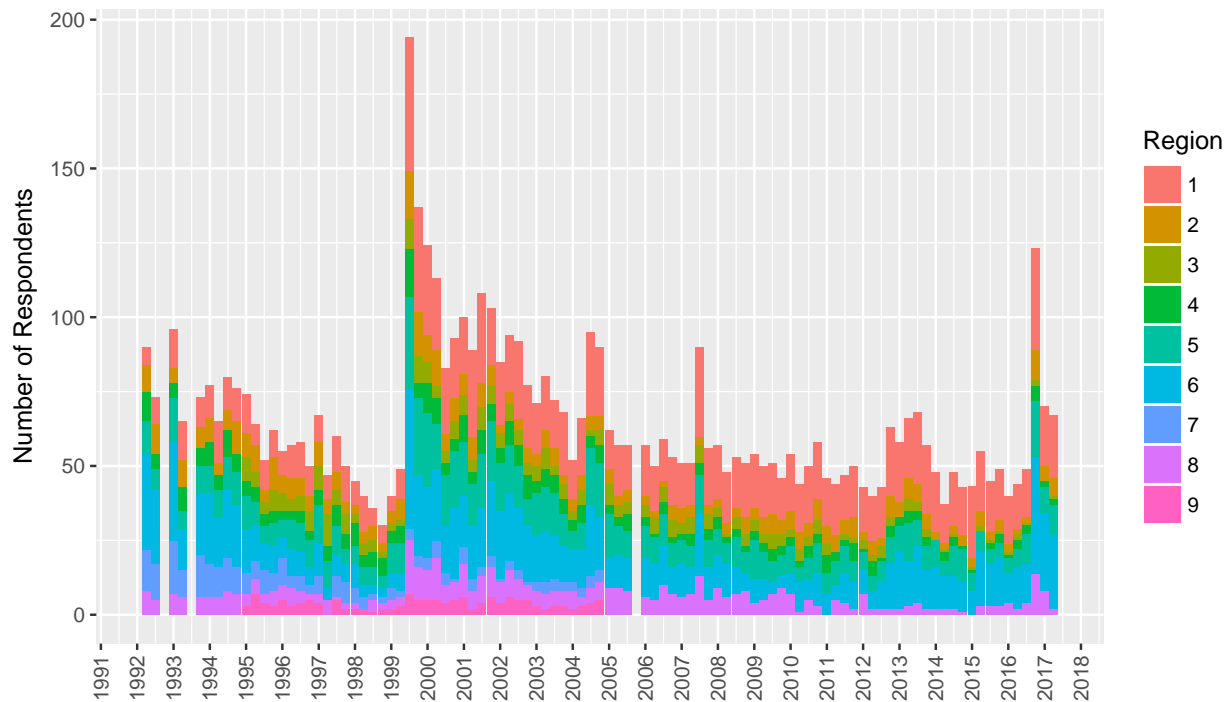


Figure 44: Number of Motor Vehicle respondents by region (1992Q2-2017Q2)

based on a new 4-part firm classification. Because there is only one subsector in the Motor surveys, the 2-step weighting procedure is not necessary (i.e. the subsector series are identical).

4.3.1.1 Published Series

In this section the published results are compared to the balances calculated with the microdata. Figure 45 illustrates the results for a selection of questions for new and used vehicles, as well as spare parts. Similar to the Retail and Wholesale surveys, there seem to have been adjustment factors added at the start of the period. These adjustment factors seem to have been relatively minor and the series seem to correspond well to the published data. Figure 46 illustrates that the published and microdata series for the three main regions and are very similar. The microdata therefore seems to be correct.

4.3.1.2 New Indicators

This section compares the microdata results using the old and new firm weights. Figure 47 illustrates the comparison for New Motor Vehicles for a selection of questions and Figure 48 compares the results for a selection of the new vehicles for the three main regions. The new weights produce series that are slightly less volatile than the old series.

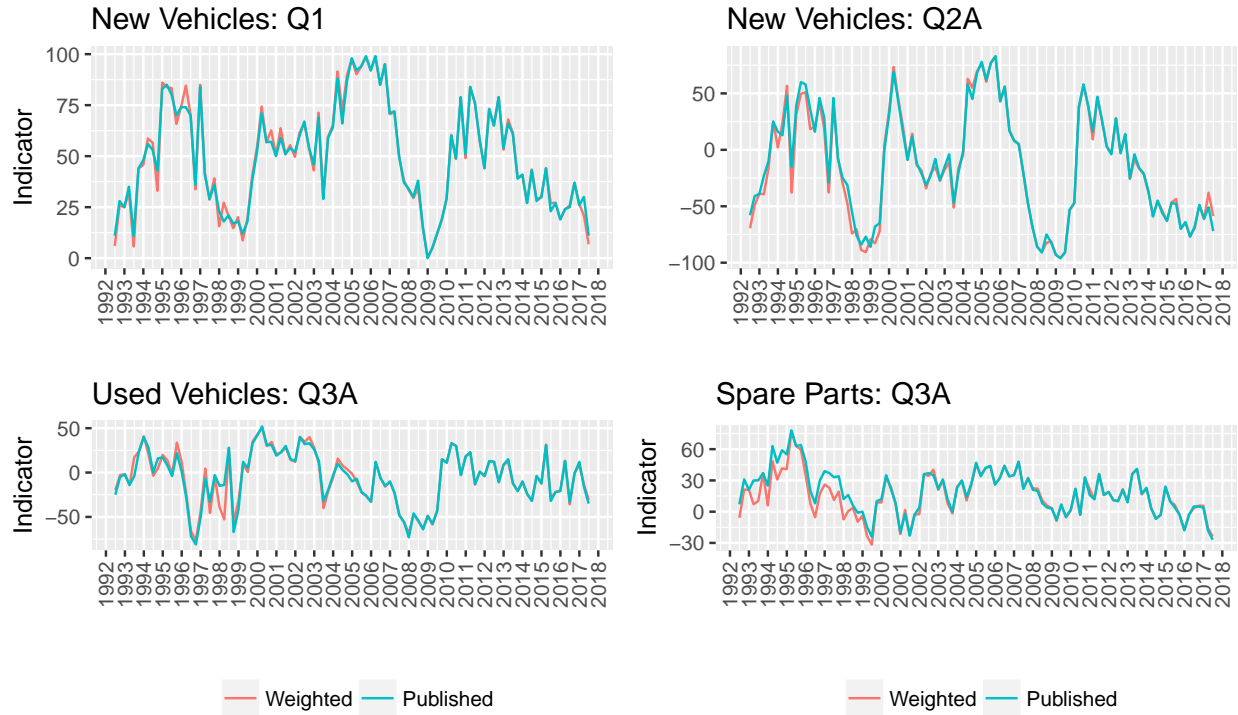


Figure 45: Motor Vehicle series compared to the published series

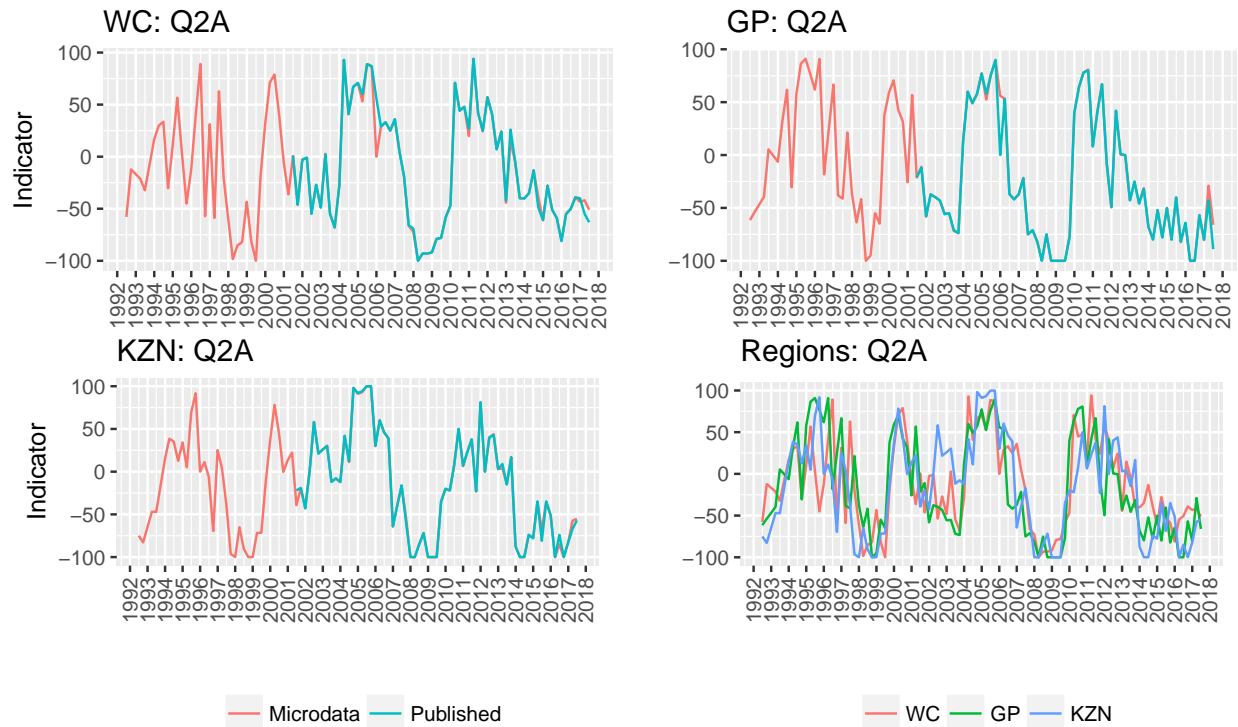


Figure 46: Motor Vehicle regional series compared to the published series

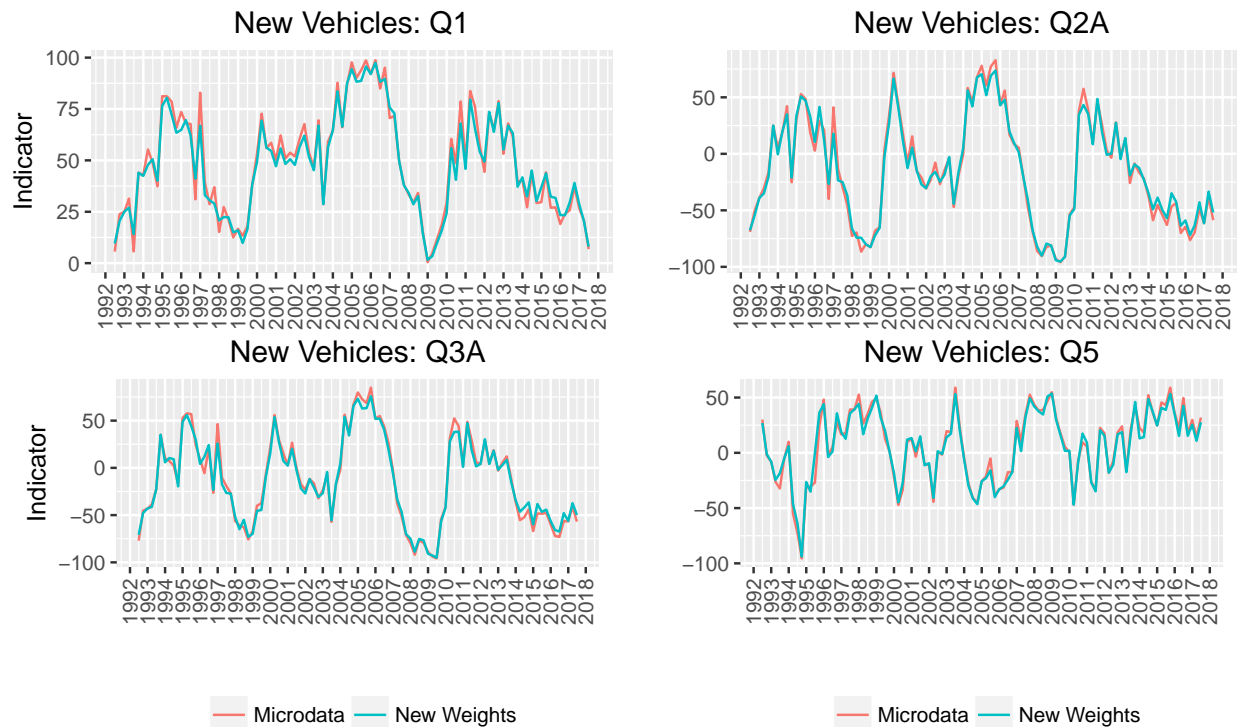


Figure 47: Motor Vehicle series using different weights

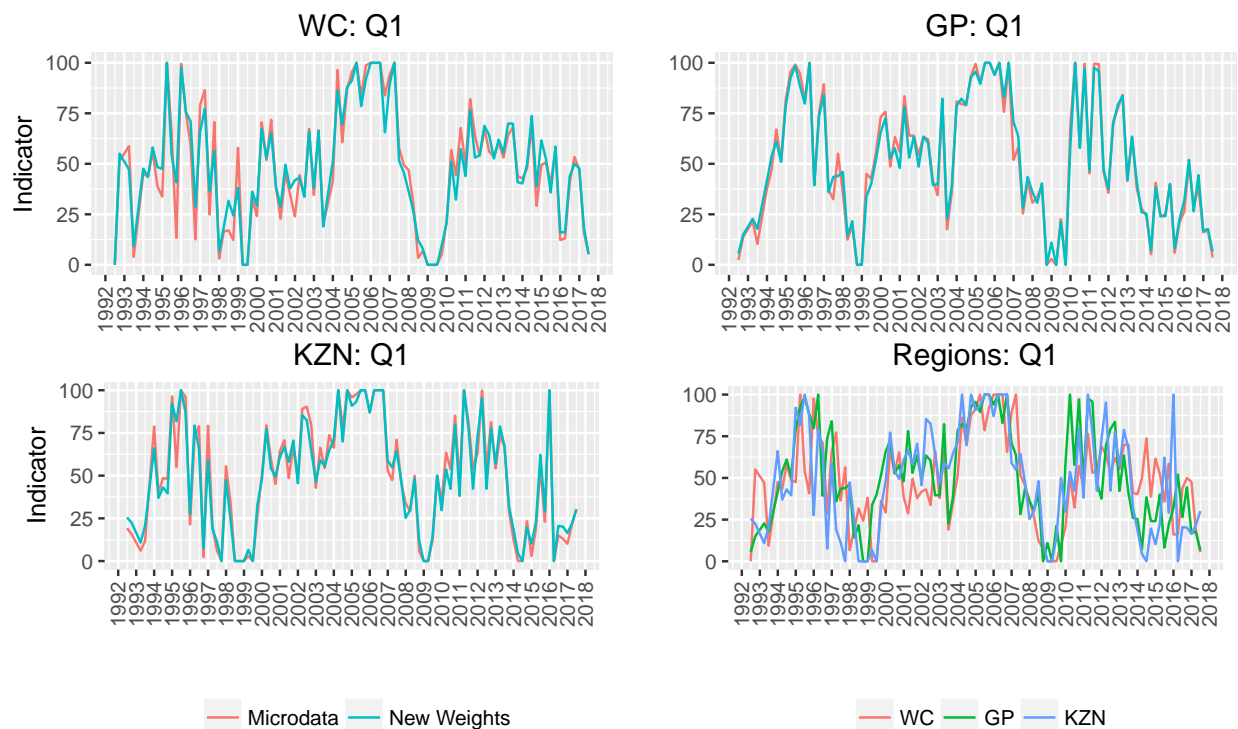


Figure 48: Motor Vehicle regions using different weights

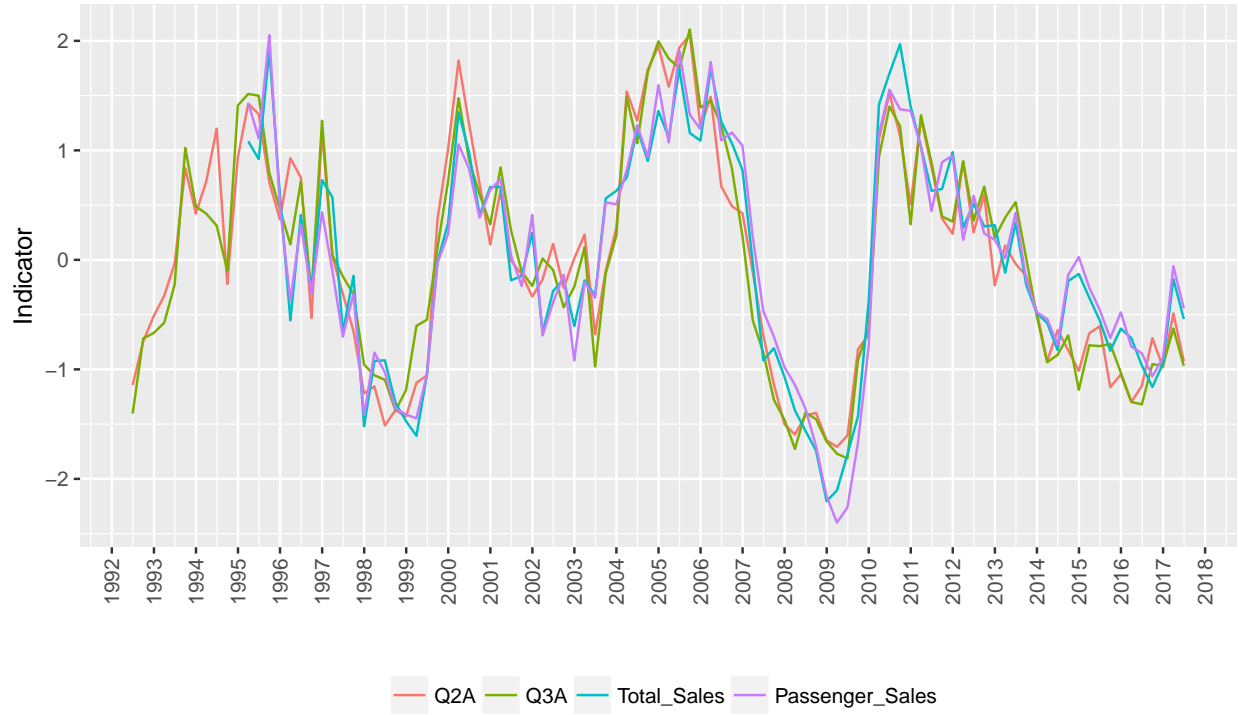


Figure 49: New Motor Vehicle sales compared to the reference series

4.3.2 Reference Series

This section compares the series to their respective reference series. The reference series that we use in this case are the NAAMSA total and passenger car sales. Figure 49 compares the balance series for total New Vehicle sales to the NAAMSA total and passenger car sales. Table 24 reports the contemporaneous correlations between the series. The correlations are very high and significant, which is confirmed by the cross-correlations illustrated in Figure 50. The new weights are produces series that are less volatile than the published data, as reported in Table 25.

Table 24: Correlations of new vehicles sales with reference series

	Q2A	Q2A_u	Q2A_n	Q3A	Q3A_u	Q3A_n
Total_Sales	0.89***	0.90***	0.90***	0.89***	0.90***	0.90***
Passenger_Sales	0.88***	0.90***	0.89***	0.89***	0.90***	0.90***

Table 25: Volatility of new motor vehicle series

	Q2A	Q2A_u	Q2A_n	Q3A	Q3A_u	Q3A_n
Volatility	47.50	43.13	44.92	46.15	41.53	43.35

Figure 51 compares the regional series for total New Vehicle sales to the regional NAAMSA total and passenger car sales. Table 26 reports that the correlations between these series are also high and significant. In general the new weights increase these correlations compared to the old weights.

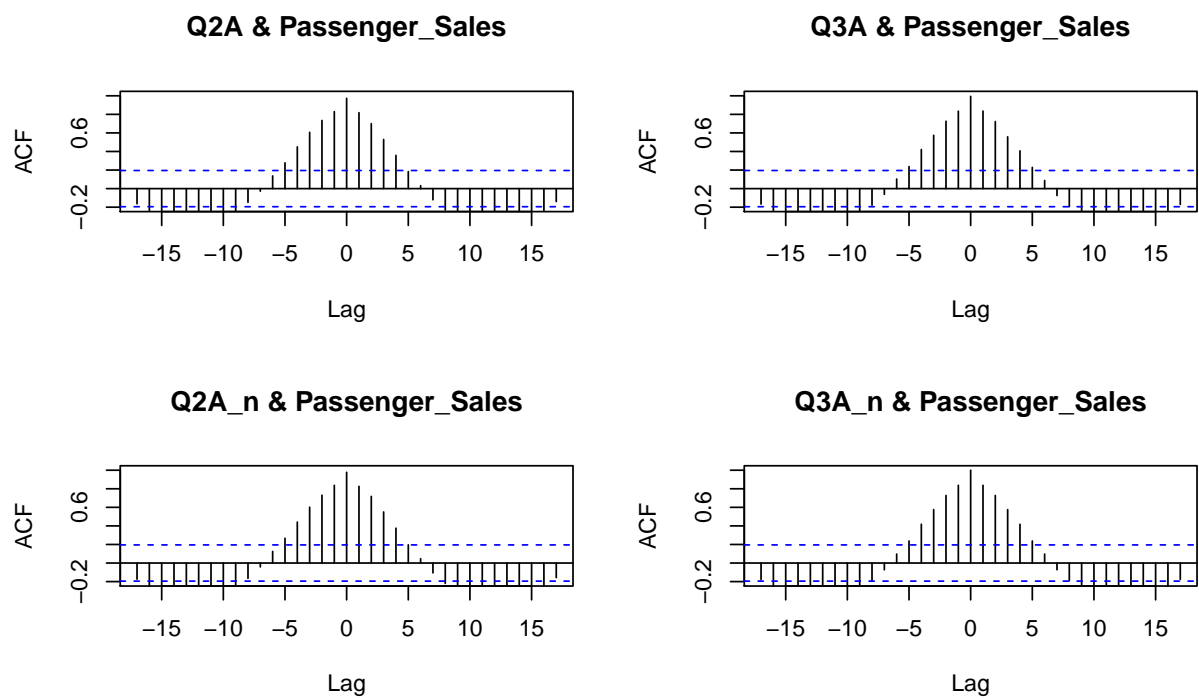


Figure 50: Cross-correlations of new vehicle sales

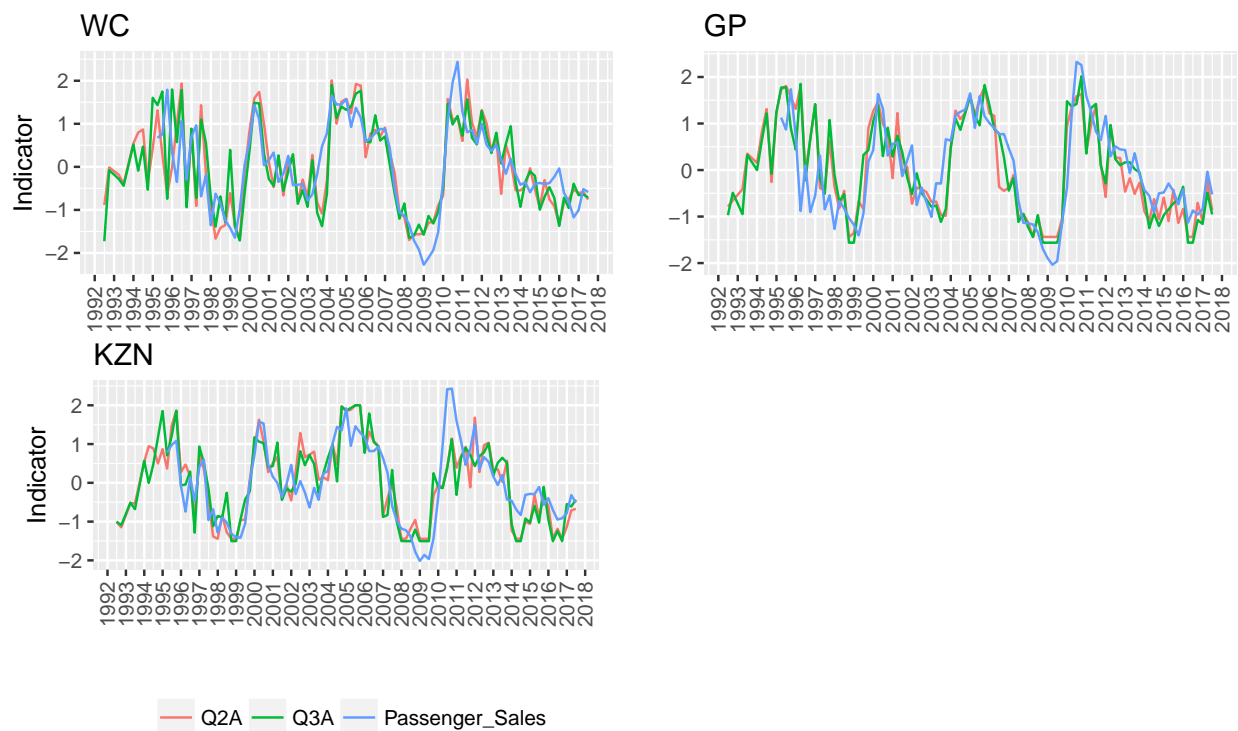


Figure 51: Motor Vehicle regions compared to the reference series

Table 26: Regional correlations with reference series

	WC_Q2A	WC_Q2A_n	WC_Q3A	WC_Q3A_n
Total Sales	0.77***	0.83***	0.75***	0.81***
Passenger Sales	0.77***	0.82***	0.75***	0.80***
	GP_Q2A	GP_Q2A_n	GP_Q3A	GP_Q3A_n
Total Sales	0.74***	0.77***	0.73***	0.76***
Passenger Sales	0.76***	0.78***	0.75***	0.78***
	KZN_Q2A	KZN_Q2A_n	KZN_Q3A	KZN_Q3A_n
Total Sales	0.80***	0.80***	0.77***	0.78***
Passenger Sales	0.80***	0.78***	0.76***	0.76***

4.4 Summary

After some corrections the microdata seems to provide series that correspond very closely to the published series. We have calculated new series and compared them to the reference series, for the retail and motor vehicles surveys. In general, the new weights and the 2-step weighting procedure produce series that are less volatile and exhibit an improved tracking record, in the sense they have higher correlations with the reference series. We still have to choose the reference series for the wholesale sector. The final series still have to be written to an Excel file, in a format that is convenient for processing. The updates can and should be fully automated.

5 The Services Survey

5.1 Collating and Cleaning the Microdata

Table 27 reports some characteristics of the Services survey microdata. The Building microdata are missing for 2005Q4. Where published series were available, we used the published values for those missing values. Where they were unavailable, we used simple linear interpolation to create complete continuous series.

Table 27: Sample characteristics

	Sample	Total Obs	Obs/Quarter	Missing Quarters
Services Survey	2005Q2-2017Q2	14155	294.9	2005Q4

Table 28 reports the subsector codes that were used to split the sample. The subsectors are *Hotels and Restaurants (Catering)*, *Transport and Storage*, *Real Estate*, *Business Services*, and *Community Services*. Figure 52 illustrates the number of respondents to the Services survey, by subsector. Because the number of respondents for Community and Personal Services are so low, we do not calculate separate balance statistics for this subsector.

5.2 The Survey Indicators

Similar to the Building survey, the Services sector survey responses do not receive weights (i.e. responses are weighted equally). The results are similar when the firm size weights are used. In order to be consistent throughout all the BER surveys, I would suggest that in future research we should create a weighted version of the Services indicators. We would need to calculate subsector size weights and the firm size weights should also be converted to four categories, as we have done with the other surveys. Once we have these weights we can apply the 2-step weighting procedure. The old sector codes (e.g. 6020, 6030, 8020, 8050, 8060, 8220, 8230, 8240 and 9060) also need to be converted to the new sector codes.

Table 28: Services Sector Codes

Sector	Code
Hotels and Restaurants (Catering)	6000, 6001, 6020, 6030, 6011
Transport and Storage	7020, 7010, 7070, 7090, 7080, 7060, 7000, 7040, 7100, 7120, 7110, 7050
Real Estate	8000, 8010, 8020
Business Services	8040, 8080, 8070, 8090, 8060, 8050, 8030, 8150, 8120, 8210, 8180, 8140, 8160, 8190, 8100, 8200, 8230, 8130, 8110, 8170, 8240, 8220
Community Services	9000, 9010, 9030, 9050, 9060, 9020, 9040

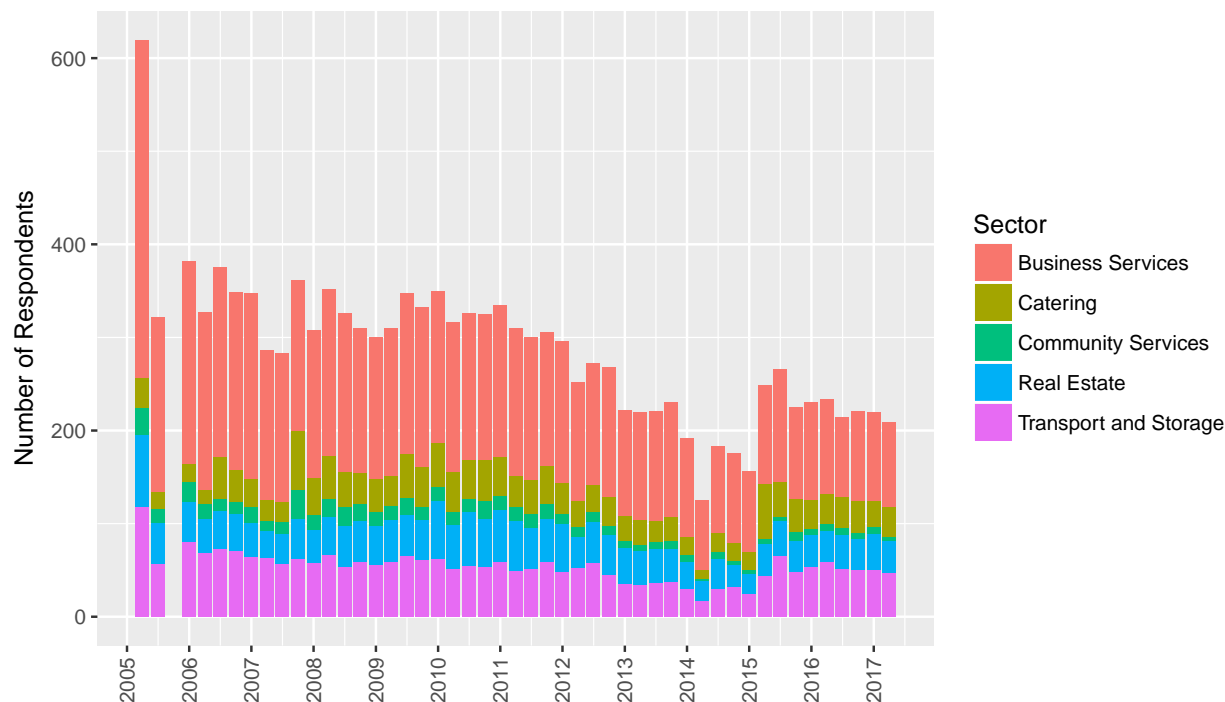


Figure 52: Number of respondents in the services sector (2005Q2-2017Q2)

5.2.1 Published Series

In this section the published results are compared to the balances calculated with the microdata. Figure 53 illustrates the Total Services series compared to the BER's published series.⁵ In many cases the published series are only available for the first few years of the sample period. Figure 54 compares the Services subsectors to the published series, and Figure 55 compares the regional total Services series compared to the published series.

Although the results do not seem to match up perfectly, the balances from the microdata seem to at least follow the same pattern as the published data. The serious exceptions are the responses to questions on constraints.⁶ These questions are weighted slightly differently from the normal balances, whereby the responses receive a weighted of 1, 0.5 and 0, instead of the standard 1, 0 and -1 quantification system. It is possible that the published series did not follow this weighting convention consistently. Be that as it may, we think that the microdata is probably correct and can be used to calculate new balance series for the Services sector.

5.3 Reference Series

We still need to choose reference series for the Services sector. Once we have chosen reference series we can evaluate the services series in terms of tracking record and volatility.

⁵As far as I understand, the BER does not really publish the results from the services surveys.

⁶Since 2015Q2, the word "skilled" has been added to the "shortage of labour constraint" question. This change in the wording of the question should be taken into account when using the time series.

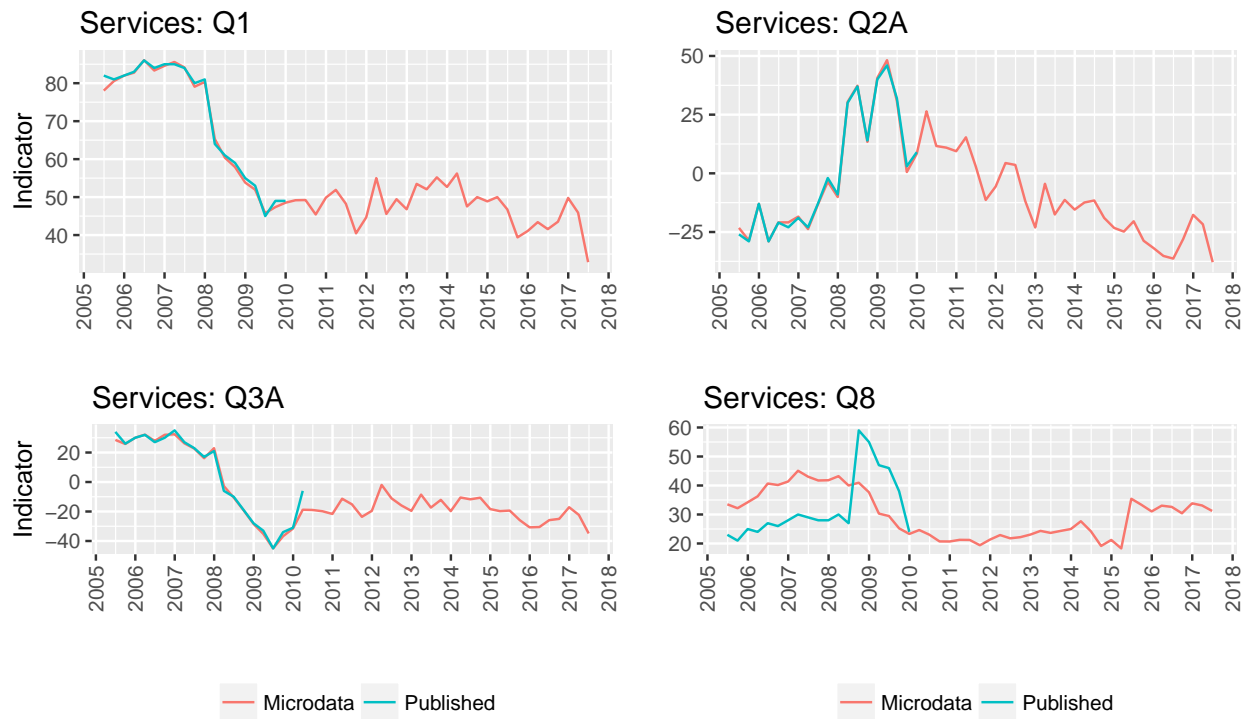


Figure 53: Total Services series compared to the published series

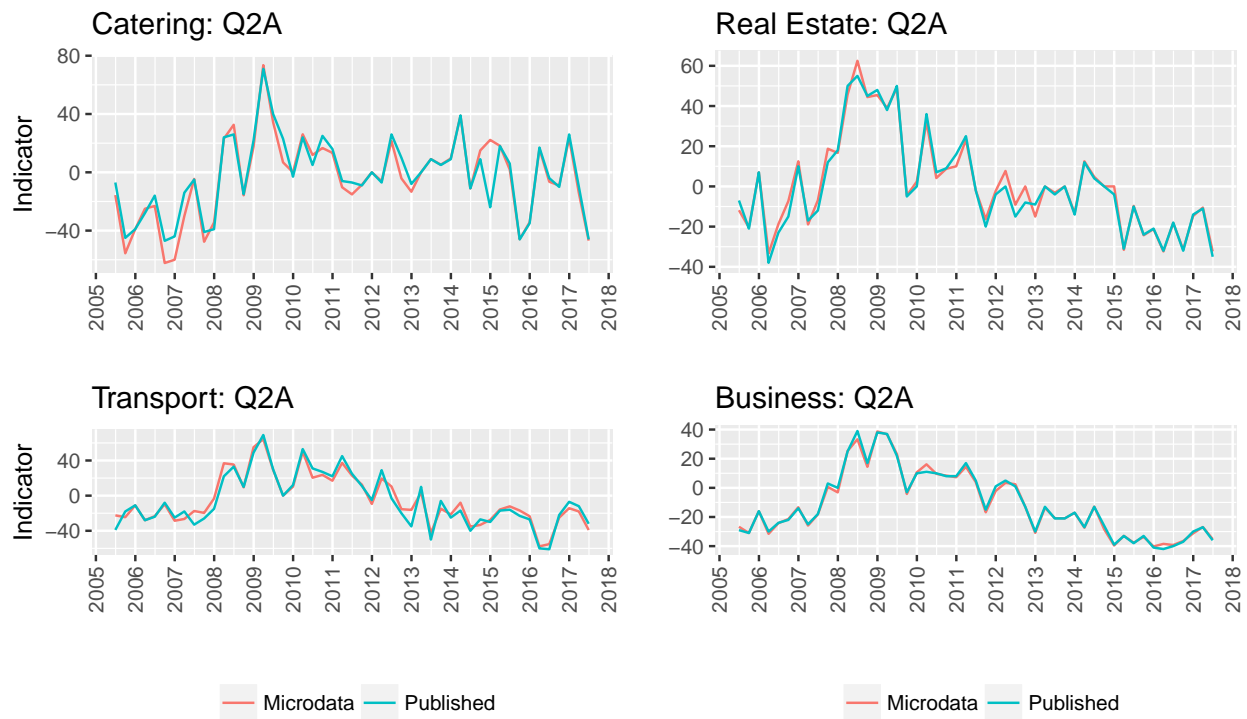


Figure 54: Services subsectors compared to the published series

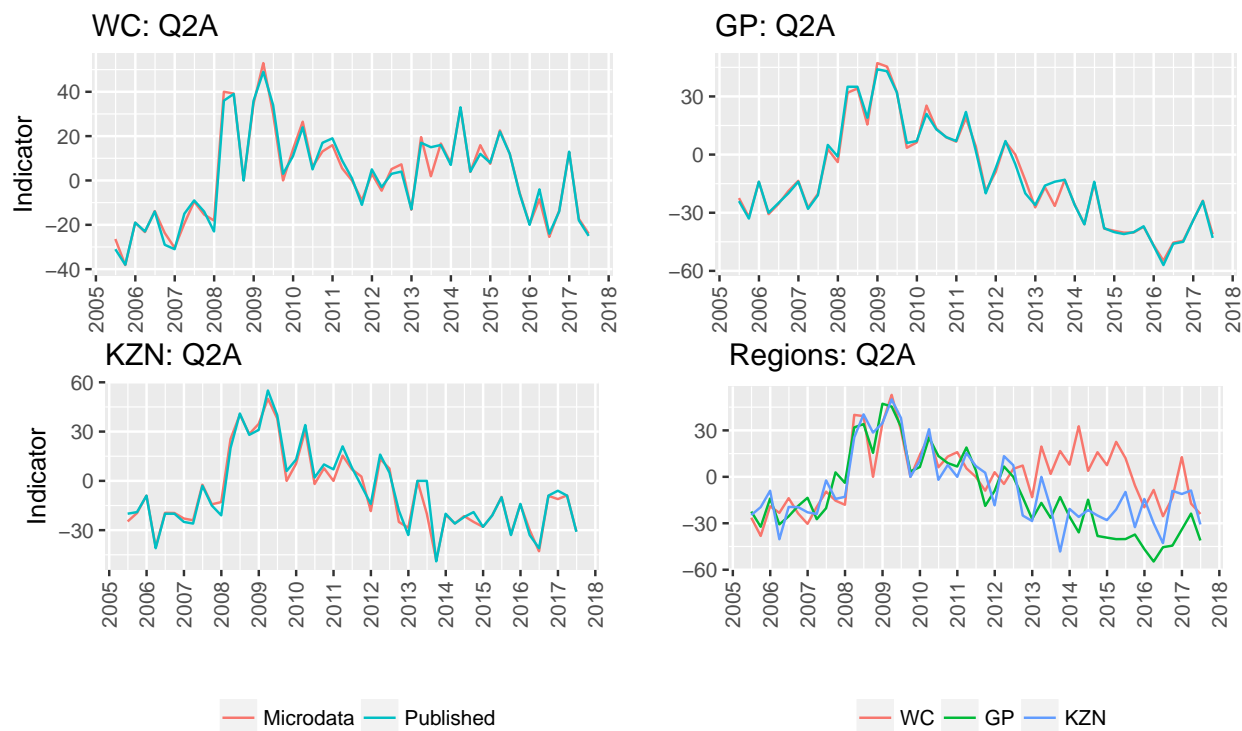


Figure 55: Services regional series compared to the published series

6 Future Research

- I would suggest that we automate the entire process (potentially in R, which is freeware), all the way from the cleaning, processing, calculations and analysis, to the presentation of the survey data. This would significantly speed up the update each quarter.
- Although the weights do not seem to make a huge difference to the final results, I would suggest that we use a consistent weighting procedure for all of the surveys. This should probably be based on the 2-step procedure recommended in the textbooks (e.g. United Nations (2015)).
- We could investigate all the series based on forward-looking questions, i.e. expected conditions. These potentially include important information and arguably reflect “confidence” more than those of current conditions. We could also combine the series, to form more stable composite indicators, as is done in most other jurisdictions, e.g. the European Commission Economic Sentiment Index.
- We could include latecomers in the following quarter, going back to the start of the period (2001).
- We could investigate the seasonality of the balance statistics. A seasonal adjustment might improve the tracking record of the series.
- We could investigate whether there are cyclical differences between different subsectors. For example, does the sub-contractor activity follow contractor activity?
- We should chain the new series to the old series, in the period before the microdata are available.
- We could look at drivers of the balance series and investigate their usefulness in terms of forecasting and nowcasting.

- We should try to present the results in a user-friendly way. For instance, when are changes significant? Should we look at moving averages or standardised series? We can compare changes to their averages, or create confidence intervals.
- We could investigate which series to focus on, perhaps using PCA. Specific series may capture most of the important information; otherwise combined composite series may be even more useful.
- We still need to find reference series for the Wholesale Survey results.
- We still need to update the weights and find reference series for the Services Surveys.

References

European Commission (2006) *European Economies: The Joint Harmonised EU Programme of Business and Consumer Surveys*. Directorate-General for Economic; Financial Affairs (Special Report No. 5).

Jolliffe, I. (2002) *Principal Component Analysis*. New York: Springer. doi: 10.2307/1270093.

Kershoff, G. (2015) ‘South Africa: The BER’s Business Tendency Surveys’, *Bureau for Economic Research*, pp. 1–19.

Organisation for Economic Co-operation and Development (2003) *Business Tendency Surveys: A Handbook*. Edited by E. Giovannini and E. Burgeat. Paris, France: OECD Publications.

United Nations (2015) *Handbook on Economic Tendency Surveys*. New York: United Nations Publications.