

# RESEARCH REPORT ON THE BER BUSINESS TENDENCY SURVEYS

Laurie Binge

This draft: November 16, 2017

## 1 Introduction

This reports sets out the BER business tendency survey results calculated with the microdata. First, we calculate the balance statistics for all of the sectors and subsectors and check that the microdata provides results that are consistent with the BER's published results. After cleaning the microdata, we calculate new series for all of the questions for all of the subsectors and the three main regions. In some cases the new series follow an explicit two-step weighting procedure. Finally, we compare the results to the reference series, to evaluate the tracking record of responses to the activity questions. In all cases the new updated series for all the questions are calculated in R and exported to Excel.

We have interpolated the 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. Duplicated 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.

Since 2016Q2, we have 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  $t - 1$ , 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 latecomers are still excluded, as we have decided to stick with this convention. The results below illustrate that the inclusion of latecomers does not influence the results substantially, with slightly changes at the subsector level. We can try to create a balanced panel, by only including those respondents that answered in consecutive quarters.

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 that were 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. All the series are then compared to the references series to evaluate their tracking record and the volatility of the alternative series.

## 2 The Building Survey

In this section we examine the Building sector surveys and results. We have used the Building survey microdata to update 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 printed and marked the questions in the old (pre-2001) Building surveys. We then matched them to the new survey questions. The old questionnaires have been uploaded to the Dropbox folder.<sup>1</sup>

There were a number of errors in the raw microdata files.<sup>2</sup> We corrected the errors in the raw microdata files relating to the questionnaires. There were also a number of errors relating to the regions codes<sup>3</sup>, which have been corrected. The corrected 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	Response Rate	Missing Quarters
Building Survey	1993Q2-2017Q2	18724	201.33	0.14	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. No high frequency data for contractors and sub-contractors is published. Figure 1 illustrates the number of respondents by subsector.

<sup>1</sup>In 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.

<sup>2</sup>In 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.

<sup>3</sup>In 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=KN, 06=GP. There was a clear change in region codes in 1995Q1, and region code before that (1993Q2-1994Q4) are different 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.

Table 2: Building sector codes

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

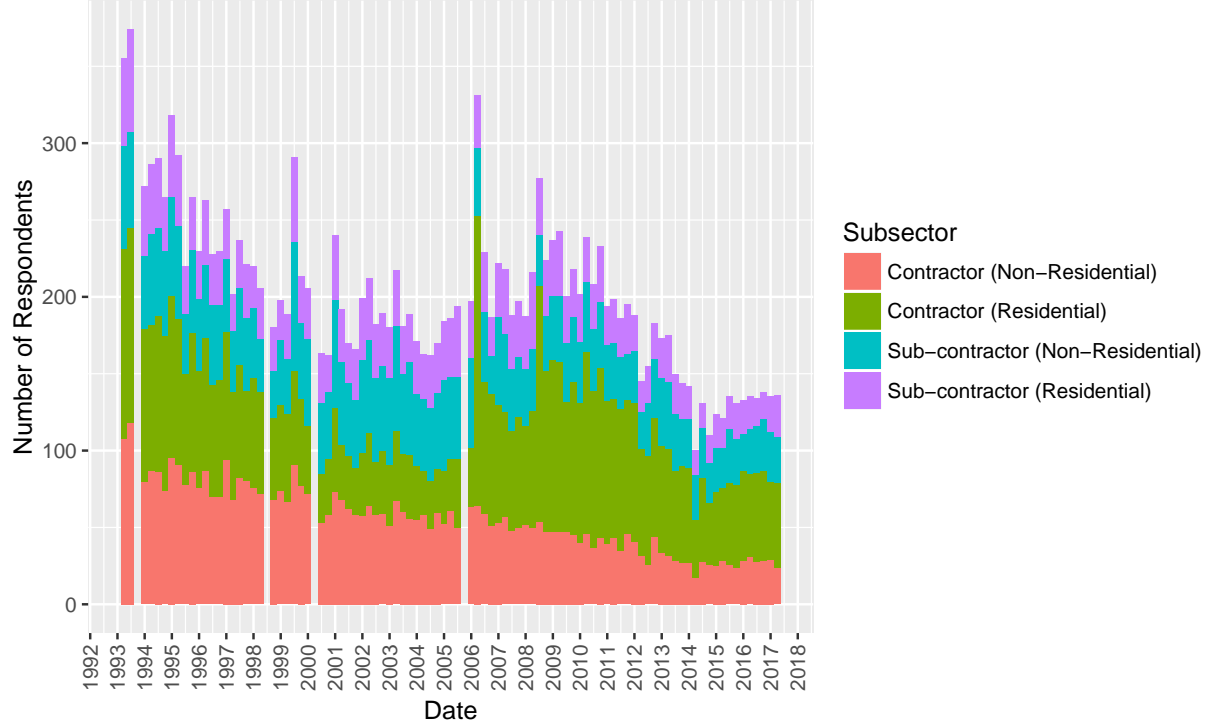


Figure 1: The number of respondents by subsector (1993Q2-2017Q2)

## 2.2 The Building Sector Indicators

Formally, one can define a  $k$ -period-ahead expectations measure of confidence ( $C_t^k$ ) 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 (usually output) 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 survey):

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

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

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

where  $a$  is determined by the preferences of the agent.

In this chapter, a distinction is made between indicators of current conditions  $C_t^k$  when  $k = 0$ , and indicators of expected conditions  $C_t^k$  when  $k = 1$ . The confidence measure for current conditions  $C_t^0$  is referred to as ‘current’, as it reflects confidence about the current quarter (in the second month of the quarter). The confidence measure for expected conditions  $C_t^1$  is referred to as ‘expected’, as it reflects confidence about the following quarter.

The BER business tendency surveys make this distinction possible by asking for separate responses relating to current and expected future conditions. The questions on current conditions (e.g. Q2A) all have the following format: “(Estimated development in current quarter) Compared with the same quarter of a year ago, are general business conditions: better, the same, or poorer?” In other words, these questions ask whether the factor under consideration in time  $t$  is better, the same, or poorer, compared with  $t - 4$ .

The forward-looking questions (e.g. Q2P) all have the following format: “(Expected development in next quarter) Compared with the same quarter of a year ago, will general business conditions be: better, the same, or poorer?” As with the questions on current conditions, these questions ask whether the factor under consideration in time  $t + 1$  is expected to be better, the same, or poorer, compared with  $t - 3$ . Responses are relative to the same quarter of the previous year, which corresponds with year-on-year growth rates.

### 2.2.1 New Indicators

The indicators (balances) for a number of groupings were calculated, as is illustrated below, without any weighting (i.e. equal weights). 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. The indicators (balances) for a number of groupings were calculated, as is illustrated below, without any weighting (i.e. equal weights). 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. There is almost no difference in the series, which is consistent with (Kershoff, 2015). Figure 2 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 missing responses correspond with those that replied, i.e. the so-called missing at random (MAR) assumption holds (Kershoff, 2015). The results below illustrate that the inclusion of latecomers does not influence the results substantially, with only slight changes at the subsector level. We therefore decided to exclude them from all of the calculation going forward, which means that the series do not need to be revised when they are updated.

Since 2016Q2, we have 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  $t - 1$ , 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 sine 2016Q2, and are

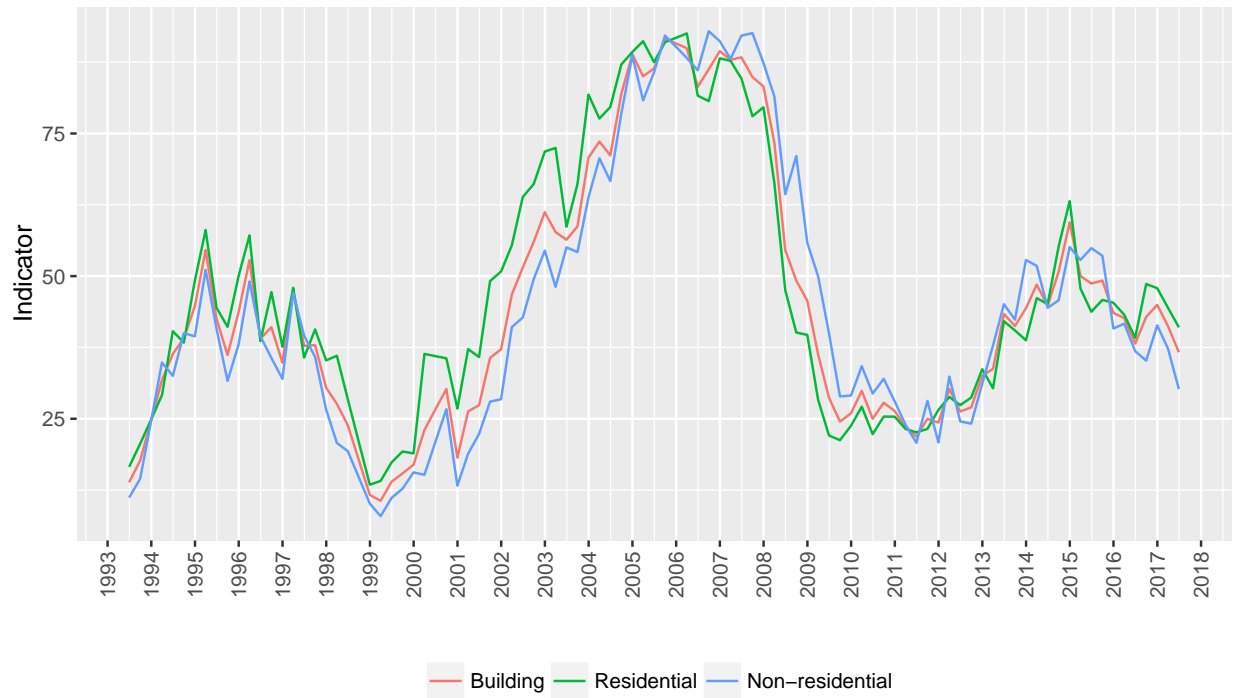


Figure 2: Building results for Q1 (confidence) in the new categories

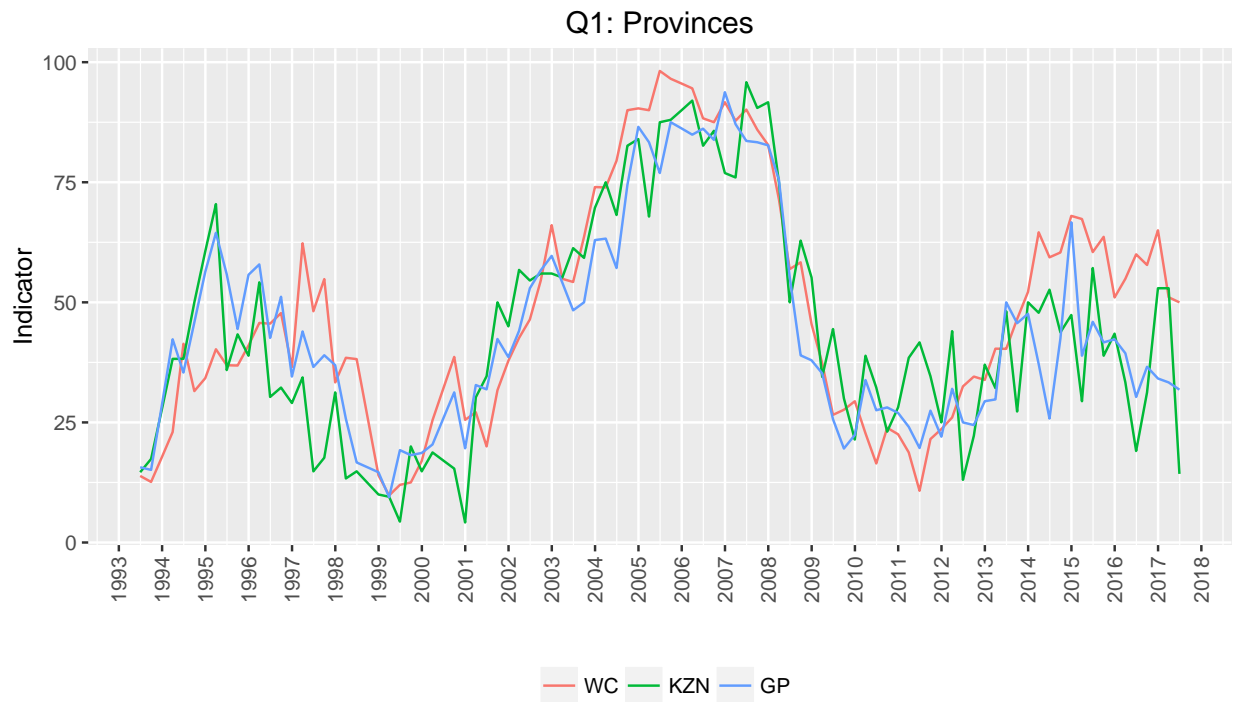


Figure 3: Building results for Q1 (confidence) for the main provinces

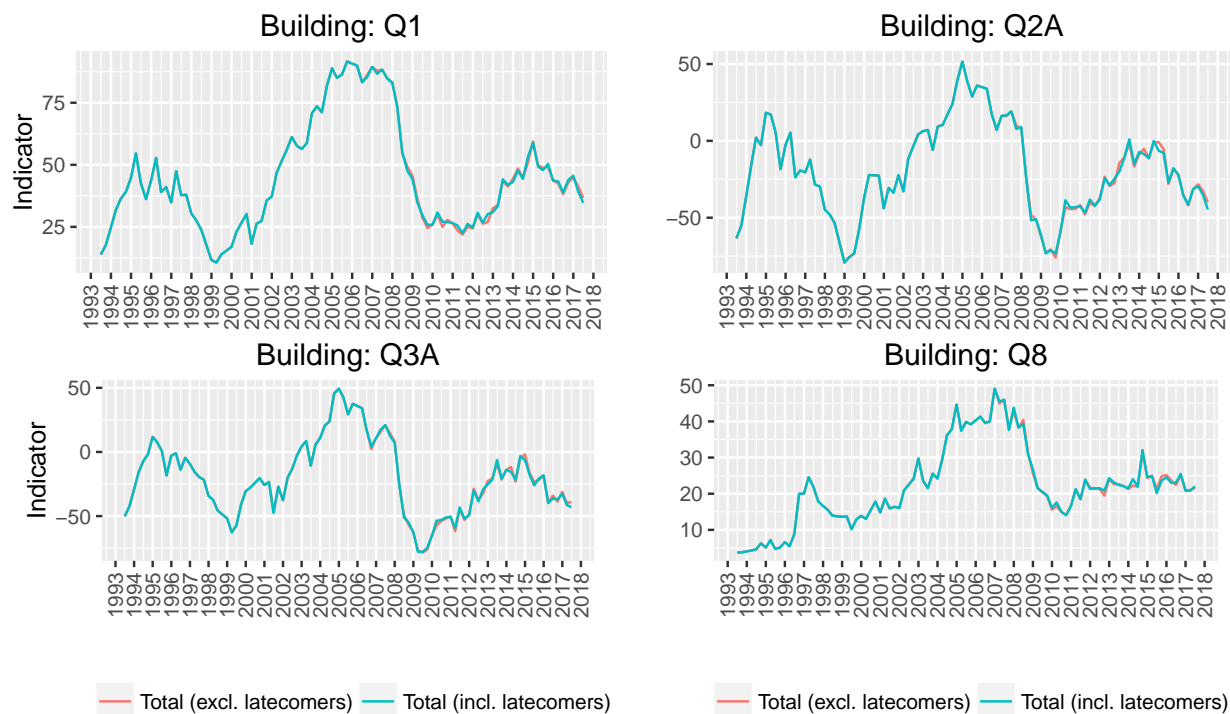


Figure 4: Building results for Q1 (confidence) - including and excluding latecomers

therefore included in the comparisons below. The latecomers are still excluded, as we have decided to stick with this convention. We could also try to do this going backwards 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).<sup>4</sup> 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. We therefore decided that the microdata could be used to calculate the new balance statistics for all the questions, as well as for the updates.

## 2.4 Alternative Weights

The balances were calculated without applying any weights (i.e. equal weights). The results are similar when the firm size weights are applied. We should create a weighted version of the Building indicators, using the two-step procedure, once we have subsector size weights. The firm size weights can also be converted to four categories, as we have done with the Manufacturing survey. Given

<sup>4</sup>The "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.

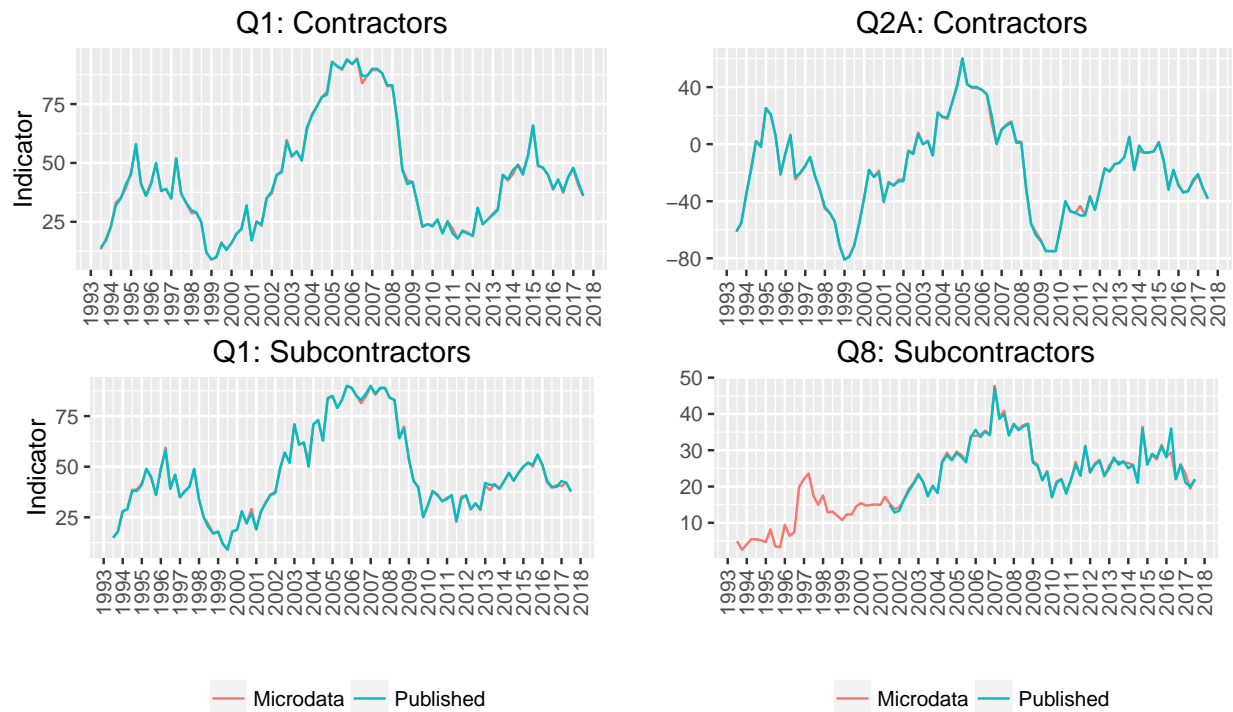


Figure 5: Microdata results compared to the published results

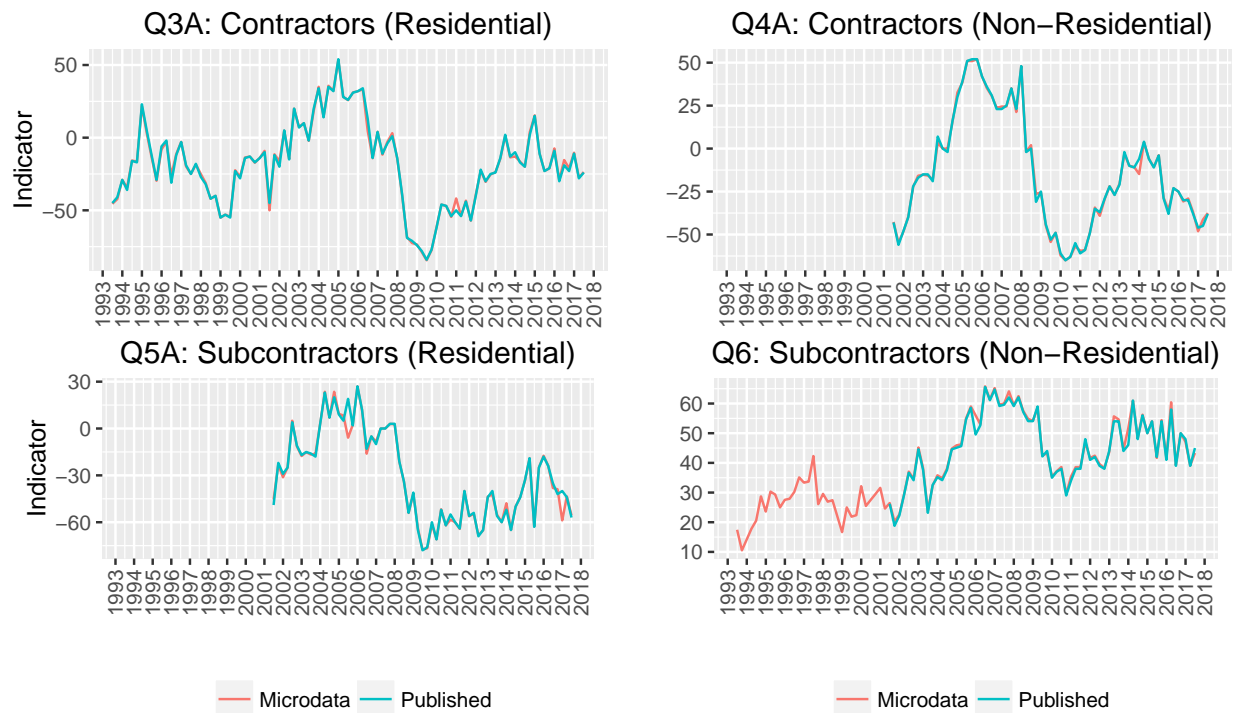


Figure 6: Microdata results compared to the published results

that the BER does not have access to the National Business Register (which represents the universe) maintained by Stats SA, it cannot calculate sample weights.

## 2.5 Reference Series

In this section the series are assessed in terms of their tracking record, i.e. their correlations to the respective reference series. The question is which reference series should be used. This is an empirical question. We compared the series graphically and calculated simple correlations.

### 2.5.1 Residential Activity

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. Building plans passed are a good leading indicator, but postponements and the stock of plans makes this data problematic as reference series (or benchmark) for the BER's building survey. 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. The SARB's data is the best match for the BER's survey, as it provides for the spread of activity over time. However, its disadvantage is that the recent data is revised extensively when more information becomes available and no provincial data is available.

Building plans passed (StatsSA) does not necessarily have a stable relationship, because planes approved do not necessarily get completed. Buildings completed (StatsSA) does not make adjustments for completions at the end of the project. Fixed investment (SARB) uses an S-curve to adjust.

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 deviation in residential activity in 2013 may therefore disappear when the SARB eventually revises its data. 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 the total. 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 at specific time points have a big impact on the growth rate and the BER does not always cover them effectively. However, the "other" category has little impact on the series over longer time periods.

An analysis of Stats SA's data shows that the share of residential buildings increased from about 60% of the total in 2001 to 80% in 2005 and 2006. Thereafter the share declined to close to 60% in 2009, but it edged up a bit since then. The change in the share of residential building activity reflects its course over this period. Residential building activity grew faster than non-residential building activity (resulting in a rising share) during the initial phase of the previous cyclical upturn (between 2001 and 2005). During the last phase of the previous cyclical upturn (2006 and 2007) and the recession (2008 and 2009), it contracted, which resulted in a declining share. In the current



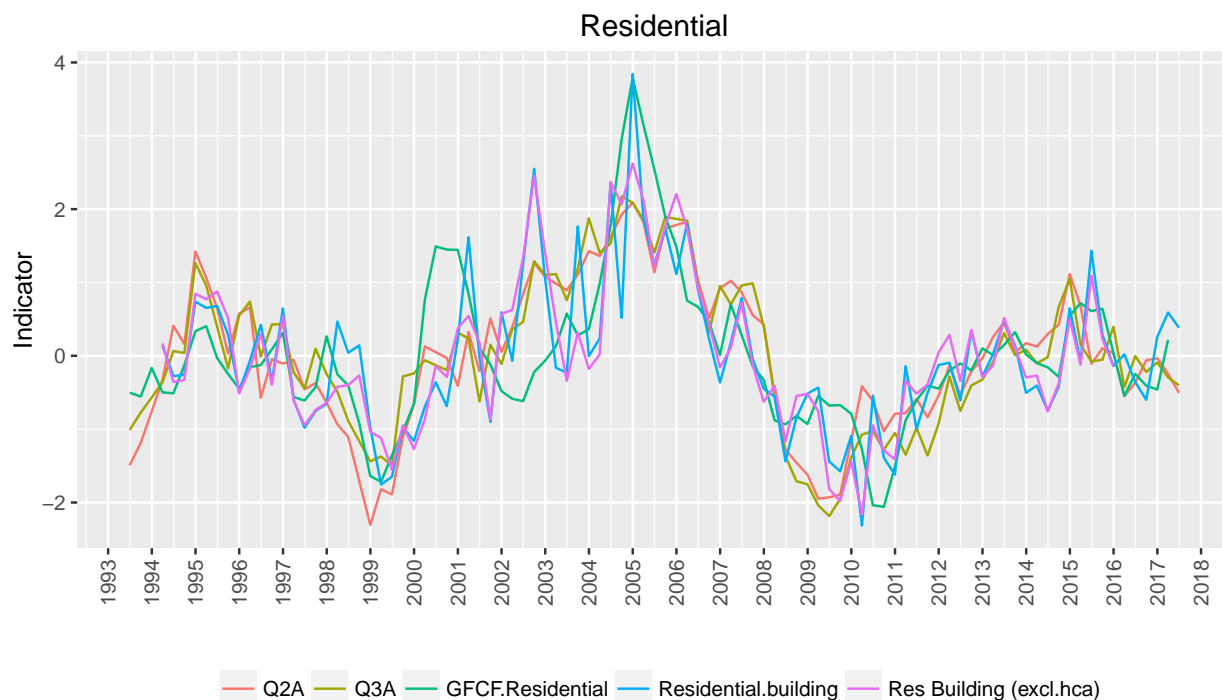


Figure 7: Residential results compared to the reference series

upswing, residential building activity underperformed relative to non-residential building activity until 2011.

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

Figure 7 illustrates the Residential series compared to the SARB and StatsSA 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 StatsSA'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 (ecl. hca)	0.77***	0.77***

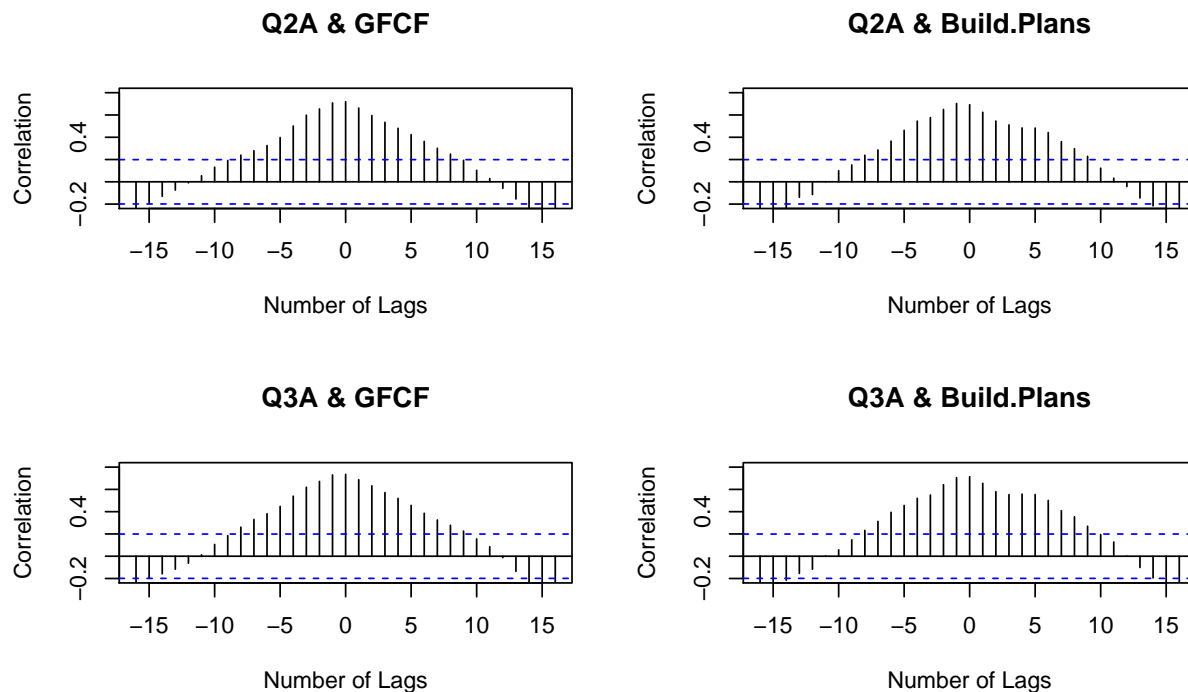


Figure 8: Residential series cross-correlations

### 2.5.2 Non-Residential Activity

Figure 9 illustrates the Non-Residential series compared to the SARB and StatsSA references series. Table 4 reports that the correlations between the survey balances and the reference series are relatively low, although the cross-correlations in Figure 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 StatsSA 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 StatsSA references series. Table 5 reports that the contemporaneous correlations between the survey balances and the reference series are relatively high. This is confirmed by the cross-correlations illustrated in Figure 12, which show again that the survey balances lead the reference series by a few quarters.

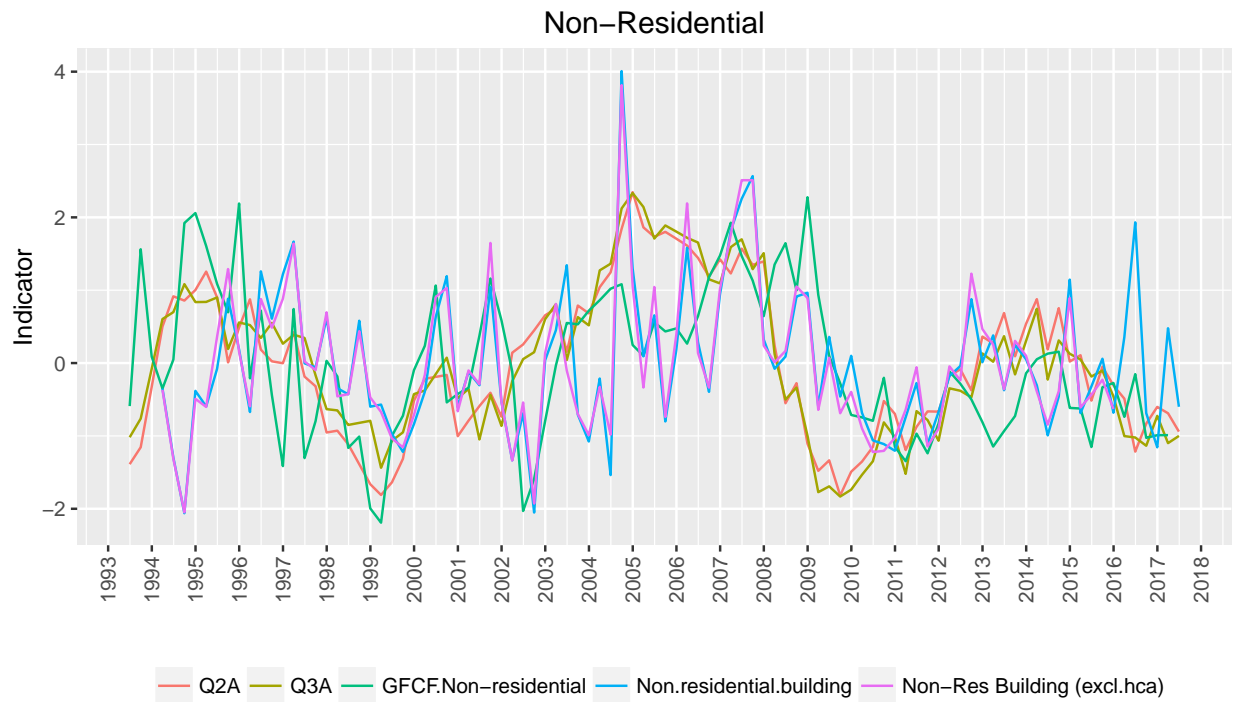


Figure 9: Non-residential results compared to the reference series

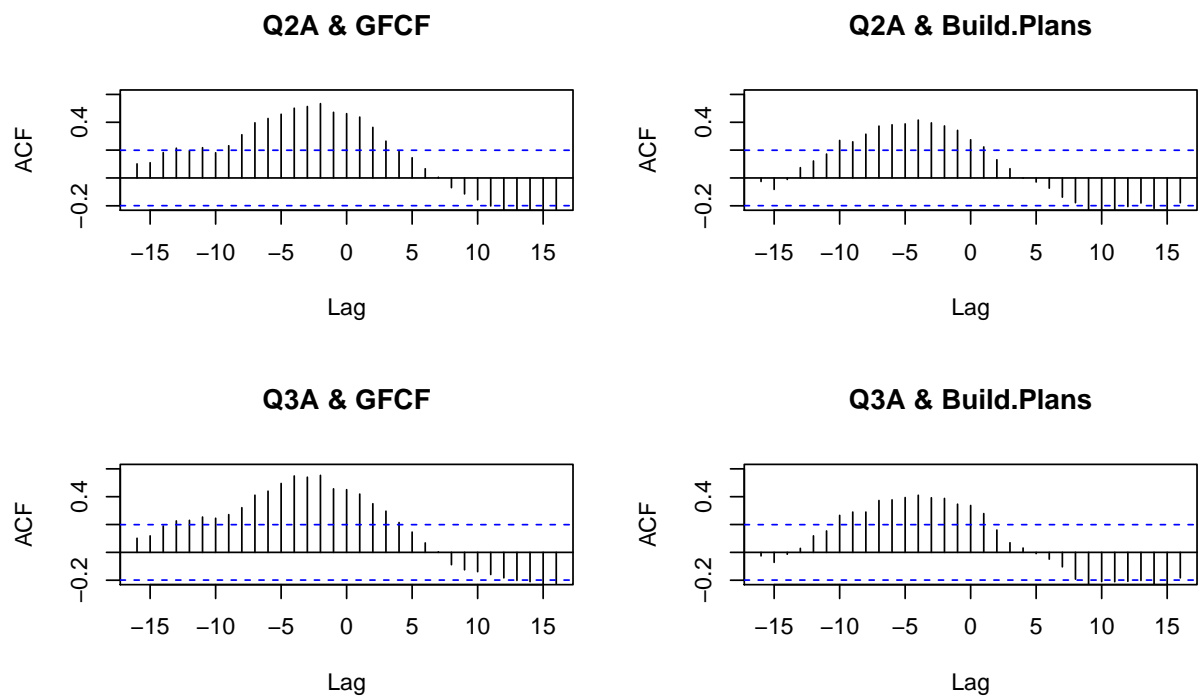


Figure 10: Non-residential series cross-correlations

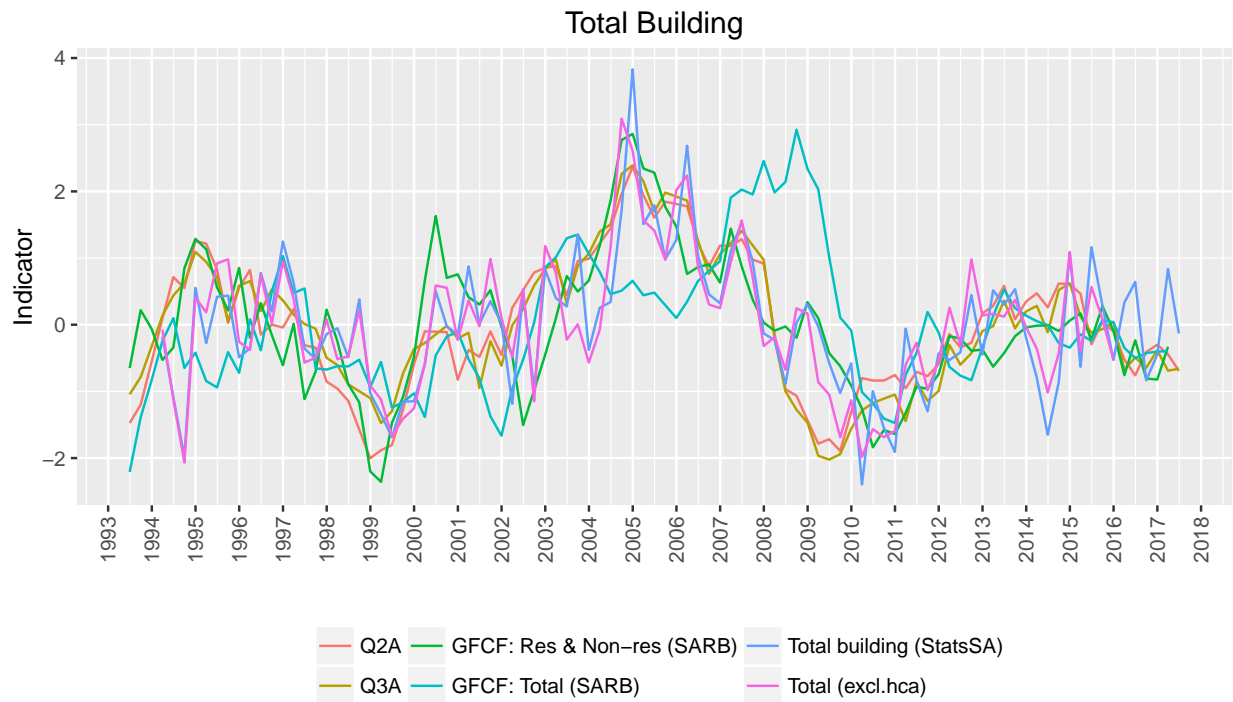


Figure 11: Total results compared to the reference series

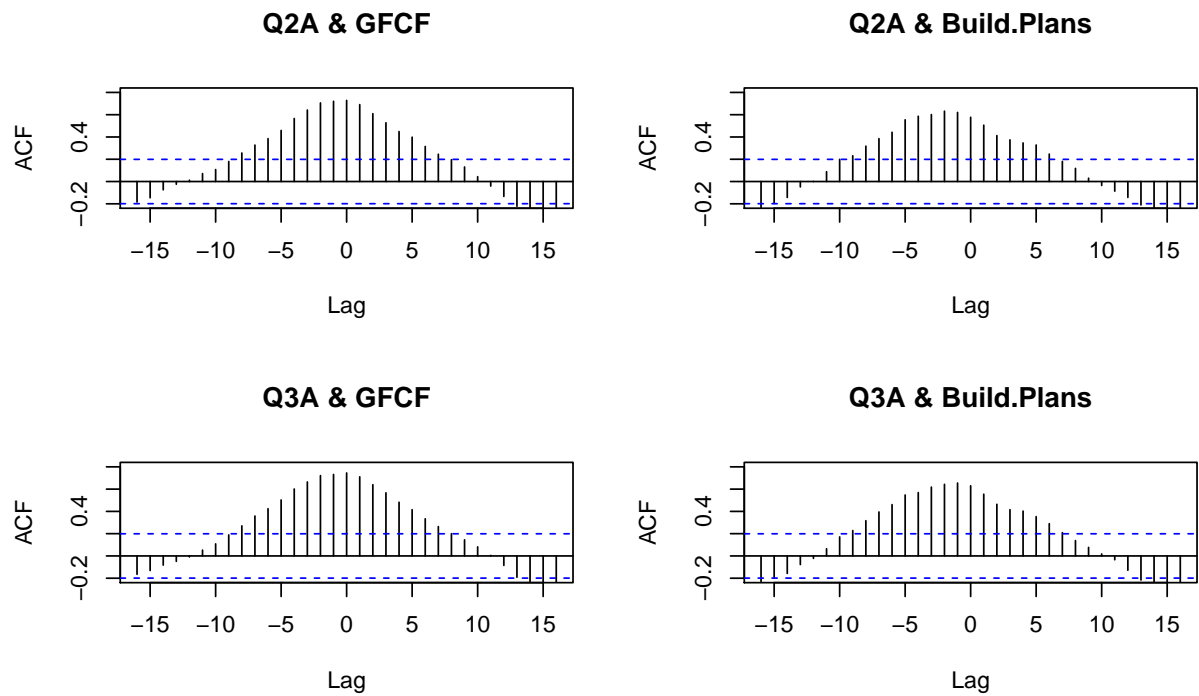


Figure 12: Total series cross-correlations

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***

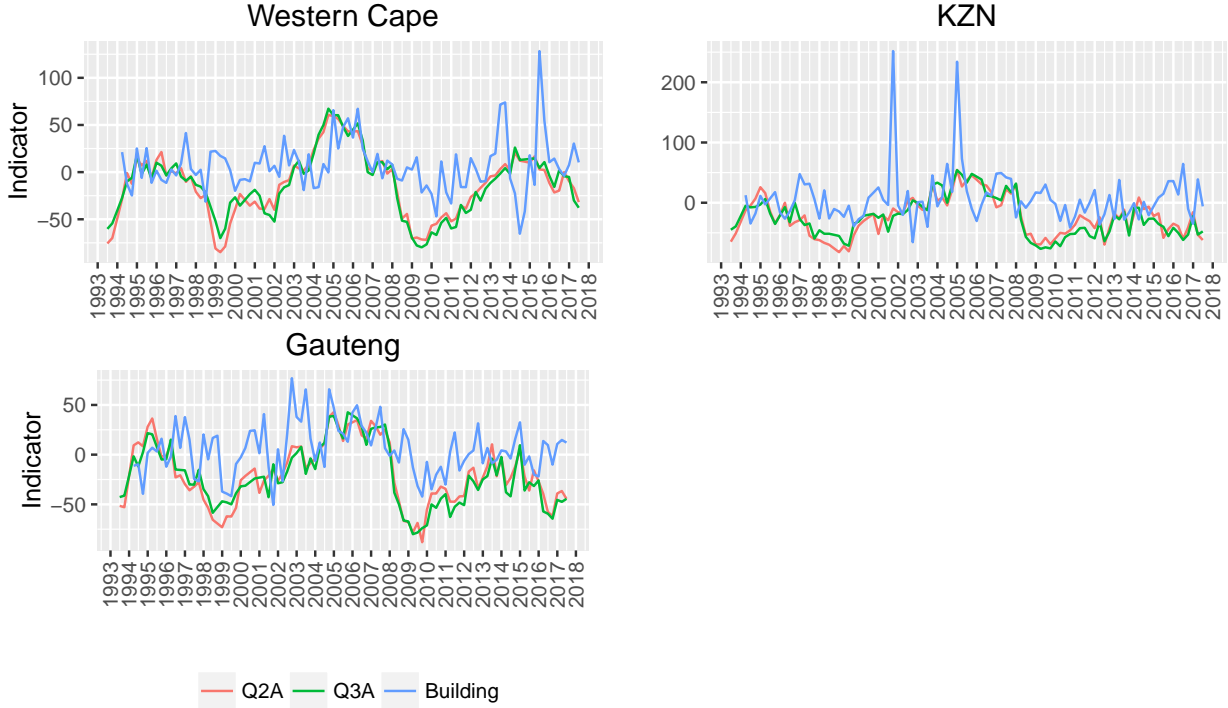


Figure 13: Regional results compared to the reference series

#### 2.5.4 Regional Activity

Figure 13 illustrates the Regional series compared to the StatsSA references series (the only one that is available). 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 StatsSA data as reference series. This means that we do not make adjustments for big projects, such as hotels and casinos, or alterations. This makes the comparison easier. The StatsSA series is also the one that is published by province, which eases that comparison and makes it consistent for all our series. For the remainder of reference series comparisons we focus on the StatsSA 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. For the Building series we should calculate weighted versions, in order to make all the series consistent.

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***

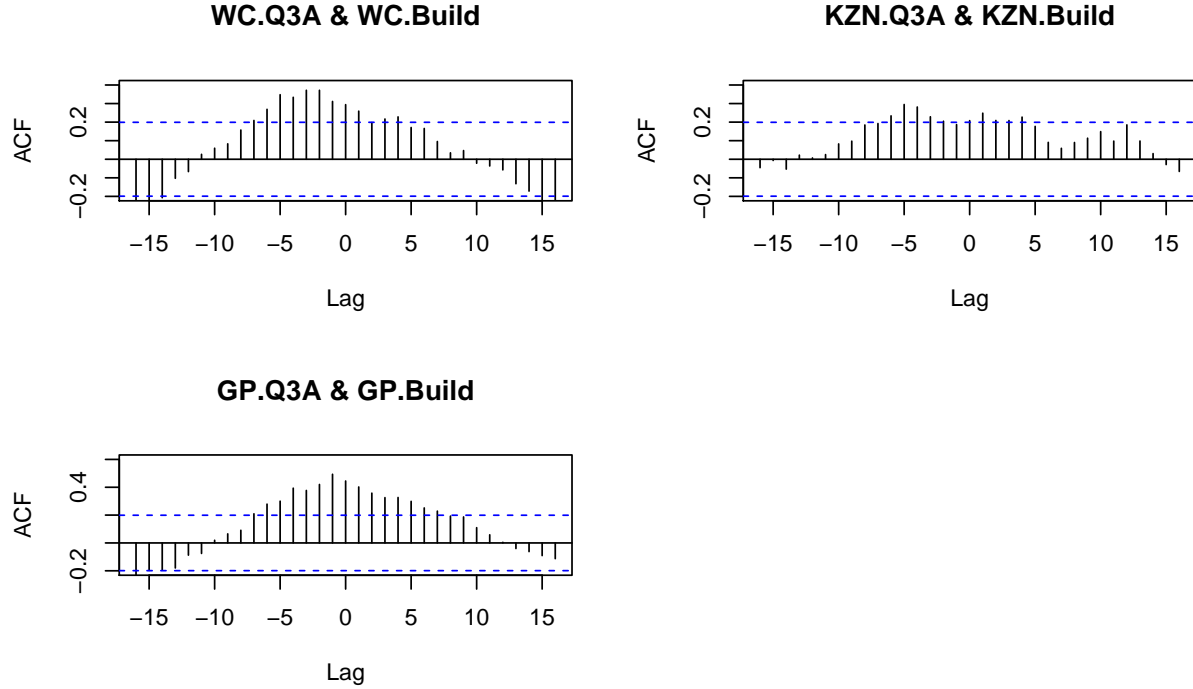


Figure 14: Regional series cross-correlations

## 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 anyway. 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.

## 2.7 Summary

So we have calculated lots of series. After some corrections the microdata seems to provide series that correspond very closely to the published series. We have decided to exclude latecomers to

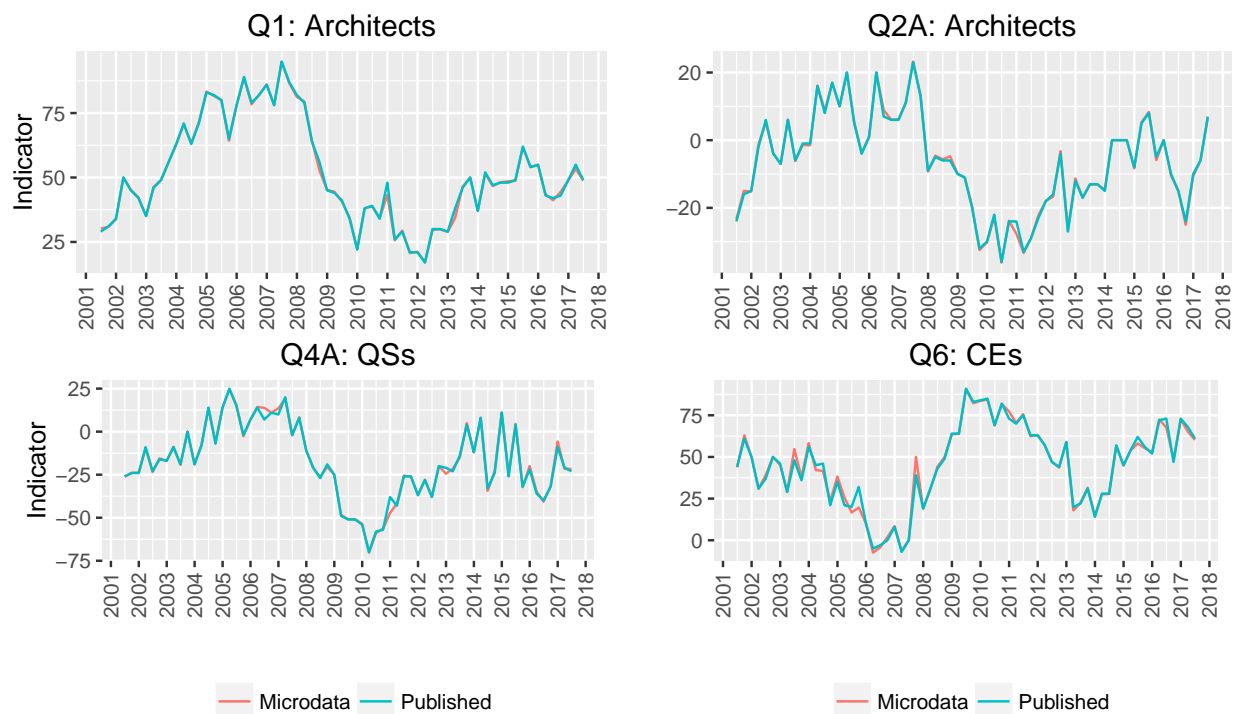


Figure 15: Architect, QS and Civil Engineer results compared to the published results

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 StatsSA series as reference series, as it is widely available for many of our series and has a relatively high correlation with the balances. We still have to try weighted versions of these indicators. 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 Building 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 printed and marked the questions in the old (pre-2001) Manufacturing surveys. We 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 these we used the codes with the highest frequencies for each id code (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	Response Rate	Missing Quarters
Manufacturing Survey	1992Q1-2017Q2	35389	364.84	0.36	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 fit into the new sector code classifications.

#### 3.2 The Manufacturing Sector Indicators

As discussed above, one can define a  $k$ -period-ahead expectations measure of confidence ( $C_t^k$ ) 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 (usually output) 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 survey):

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



Table 8: Manufacturing subsector weights

<b>New BER classification</b>	<b>Old BER classification</b>	<b>Sectors</b>
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

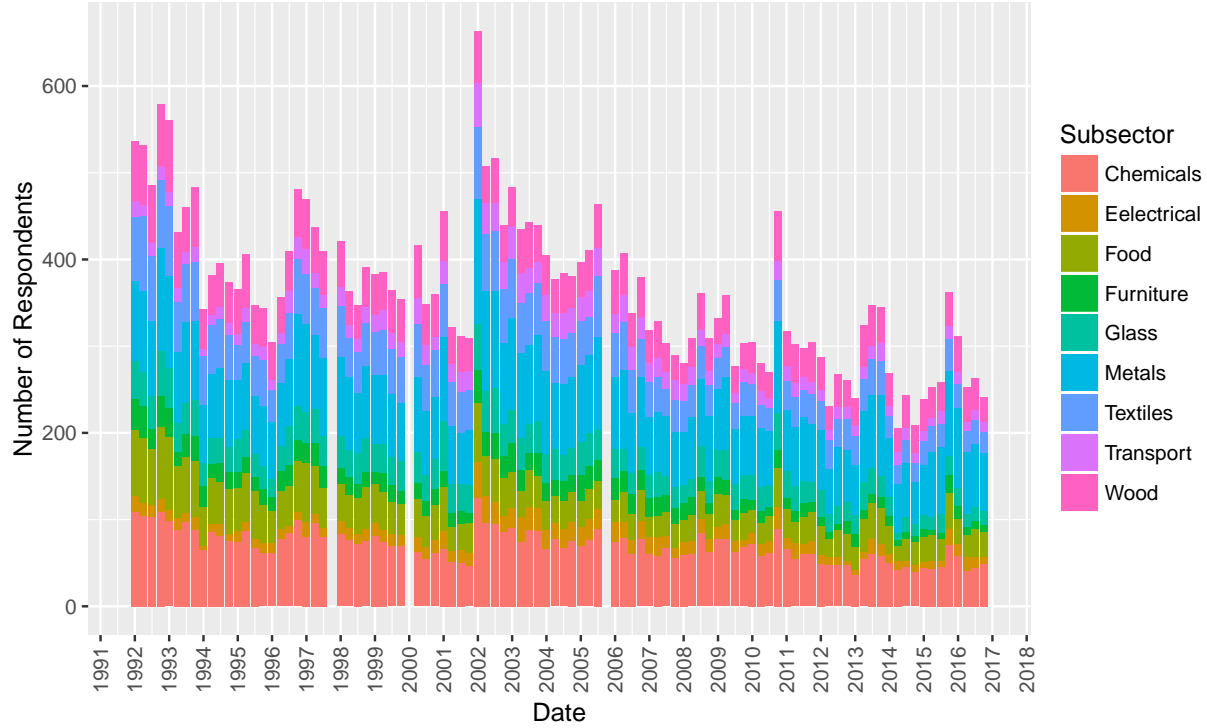


Figure 16: The number of respondents in the manufacturing sector by subsector (1992Q1-2017Q2)

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

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

where  $a$  is determined by the preferences of the agent.

In this chapter, a distinction is made between indicators of current conditions  $C_t^k$  when  $k = 0$ , and indicators of expected conditions  $C_t^k$  when  $k = 1$ . The confidence measure for current conditions  $C_t^0$  is referred to as ‘current’, as it reflects confidence about the current quarter (in the second month of the quarter). The confidence measure for expected conditions  $C_t^1$  is referred to as ‘expected’, as it reflects confidence about the following quarter.

### 3.2.1 Alternative Weighting Procedures

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. This is the weighted average of the responses, using the cleaned version of the microdata and the weights used by the BER. More formally, the cross-sectional mean of survey responses if the standard quantification system is used: ‘better’ is quantified by +1, ‘the same’ by 0, and ‘poorer’ by -1. Confidence in period  $t$  relating to current

conditions  $C_t^0$ , and confidence 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 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.

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 composition of production or sales in each subsector, as calculated by StatsSA. 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 StatsSA's and the DTT'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 an explicit two-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  $t$ . These weights are equivalent to an explicit two-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). Given that the BER does not have access to the National Business Register (which represents the universe) maintained by Stats SA, it cannot calculate sample weights.

### 3.2.2 Published Series

The first step is just 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.

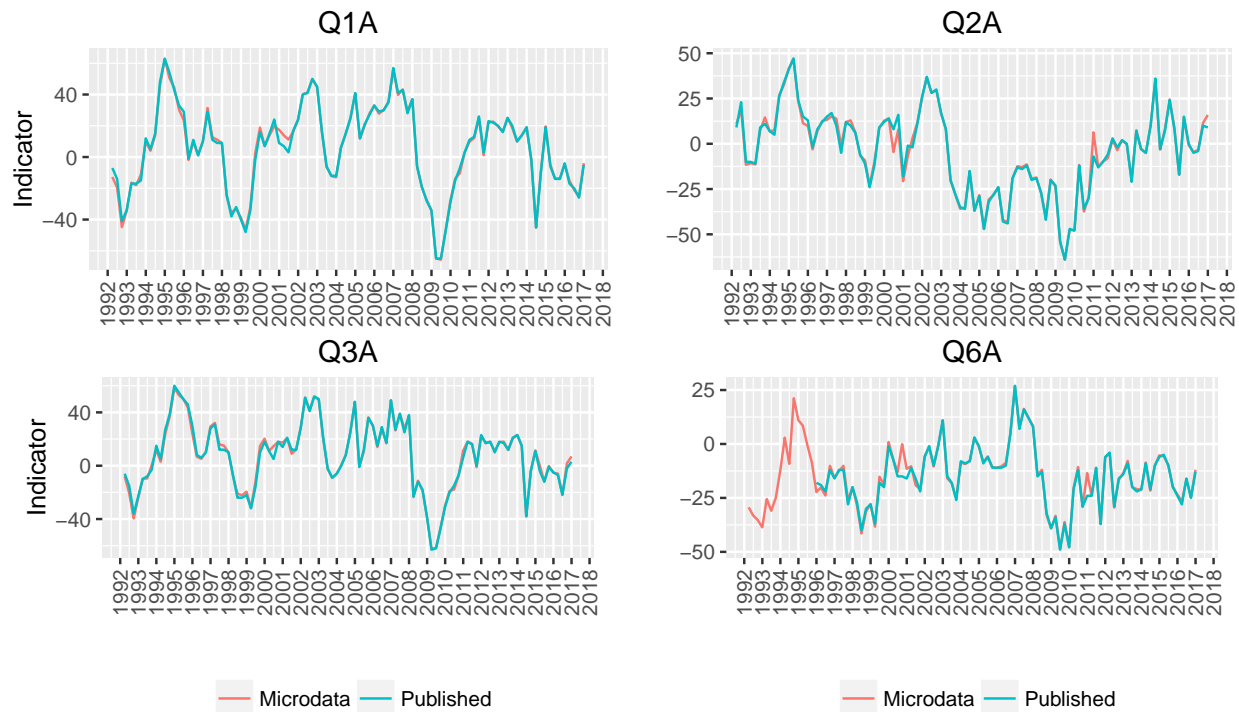


Figure 17: Manufacturing microdata results compared to the published results

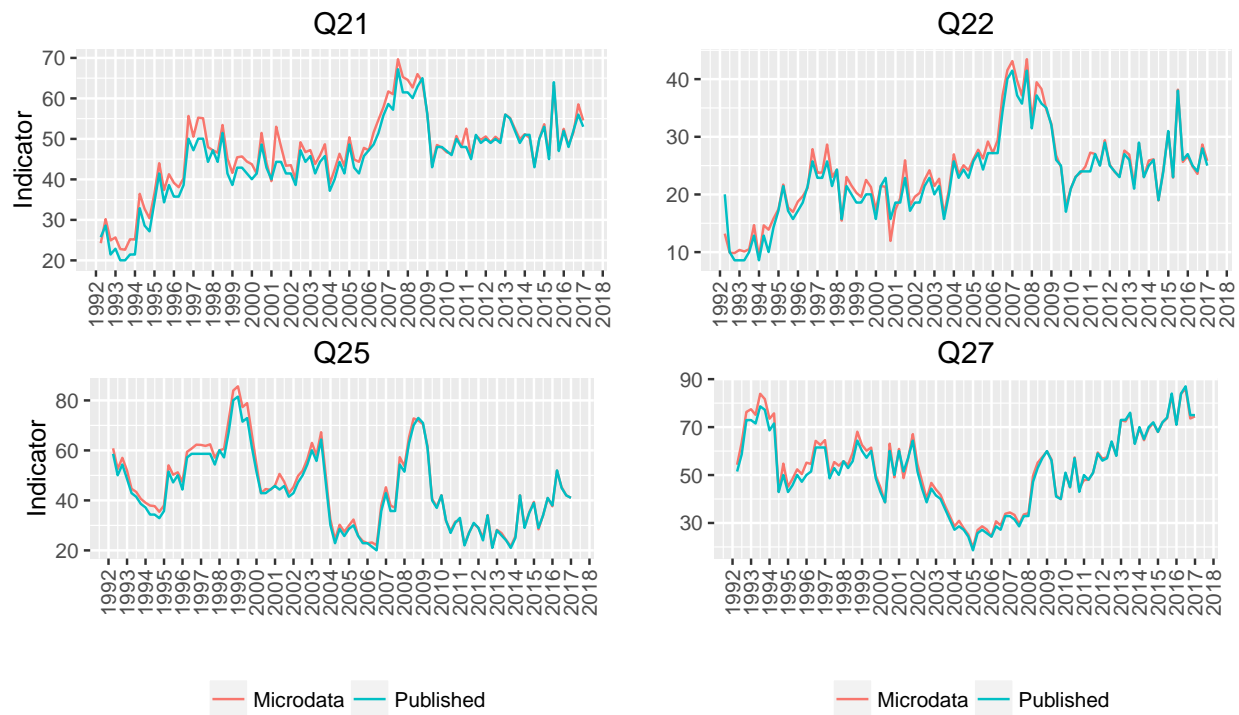


Figure 18: Manufacturing microdata results compared to the published results for constraints

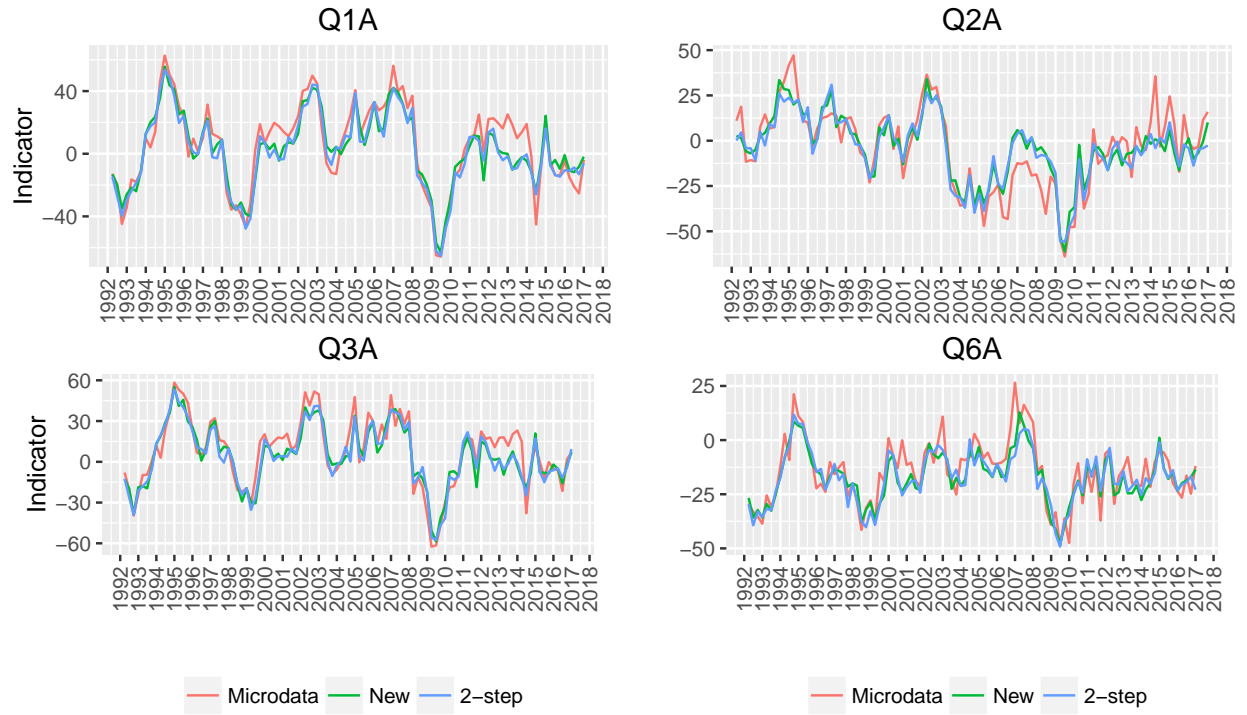


Figure 19: results

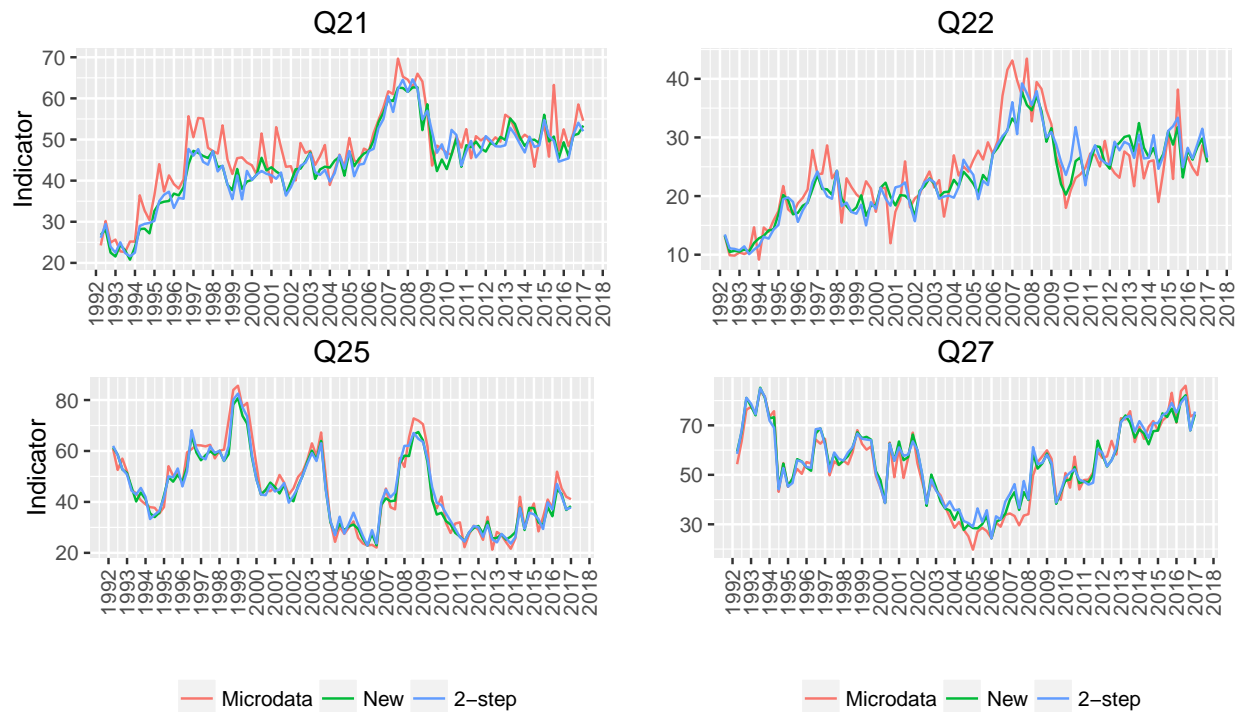


Figure 20: results

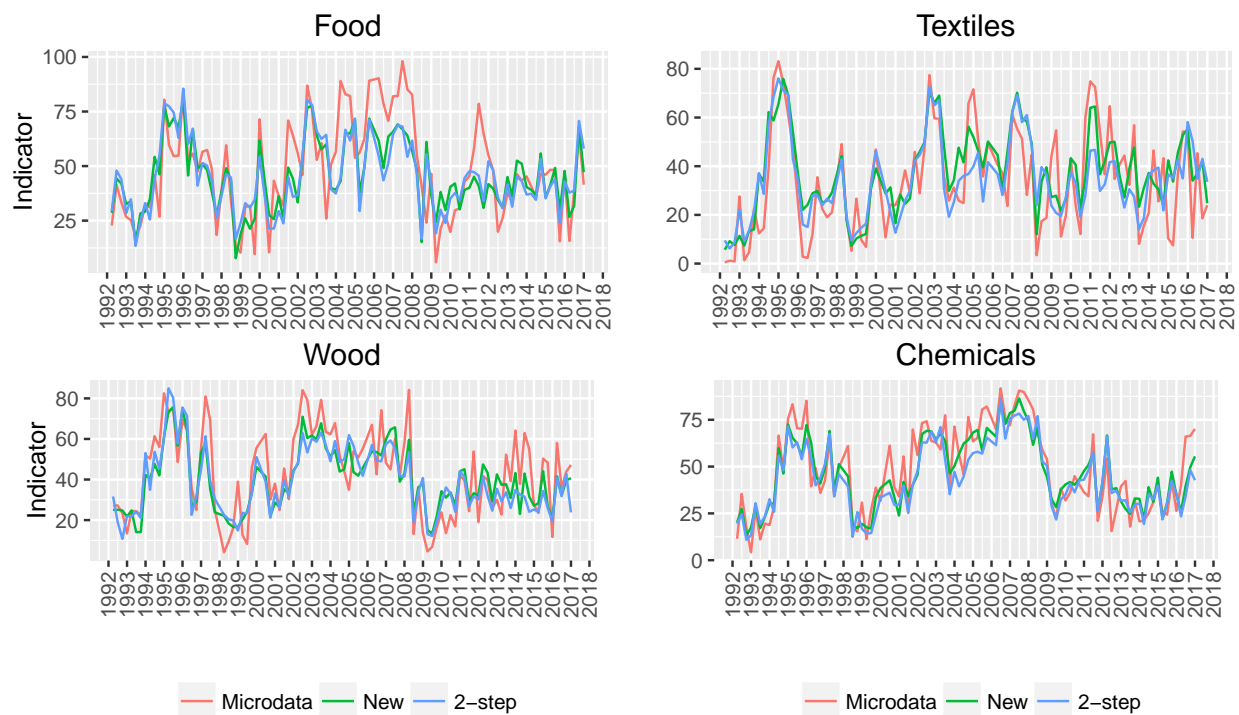


Figure 21: results

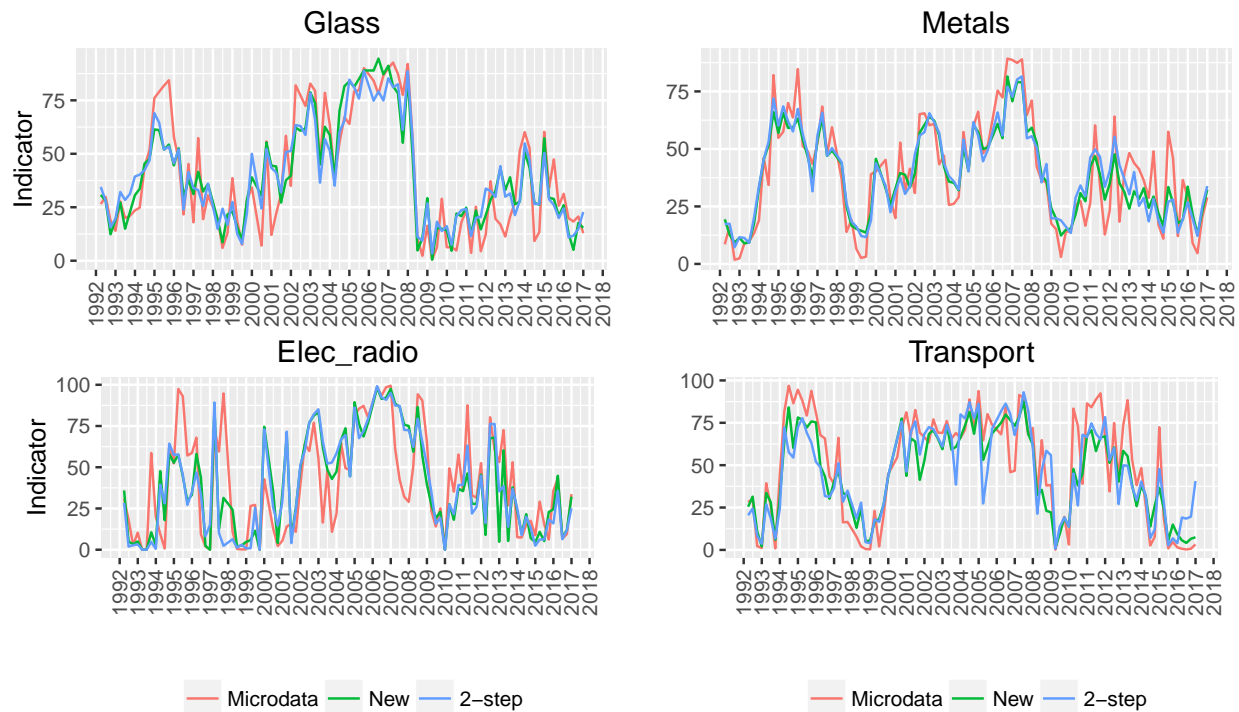


Figure 22: results

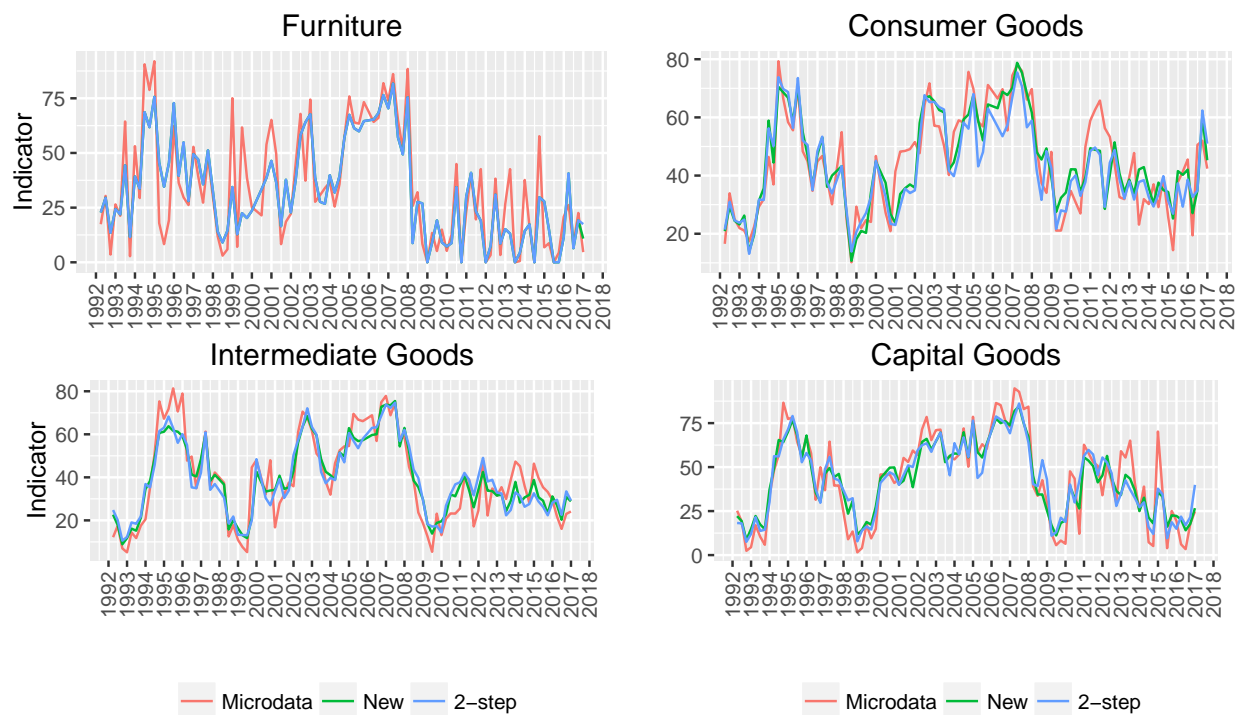


Figure 23: results



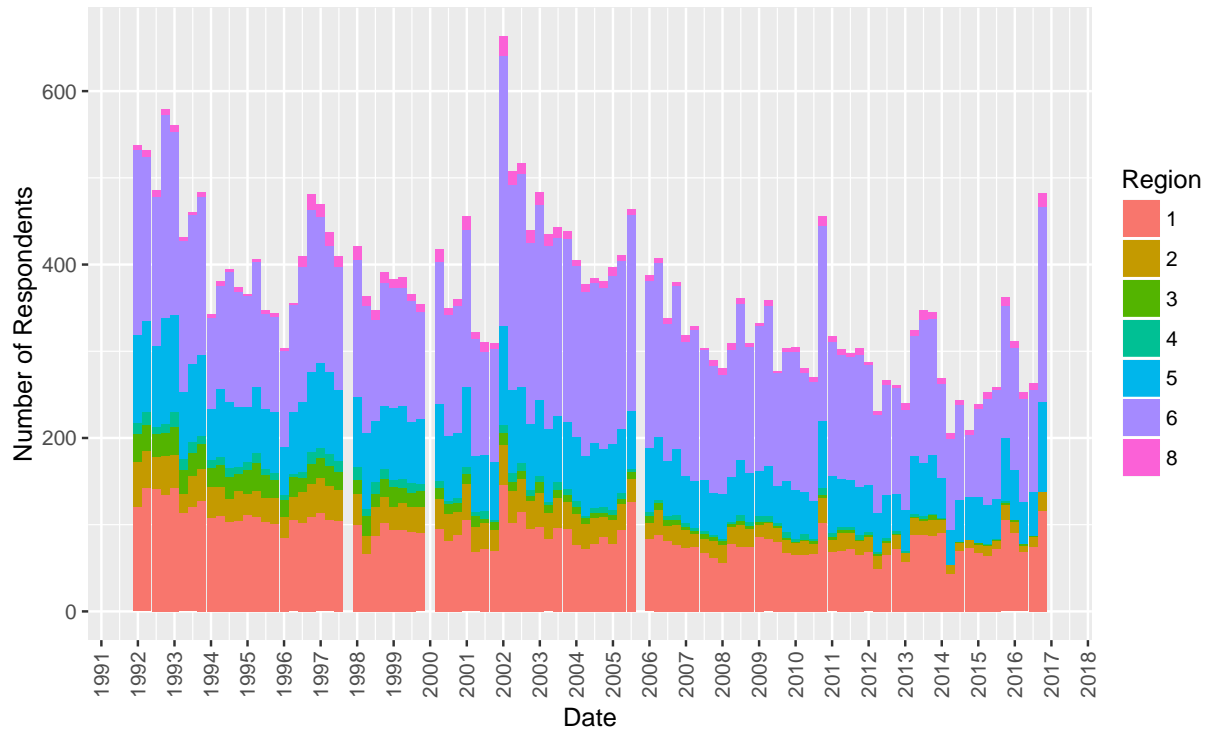


Figure 24: results

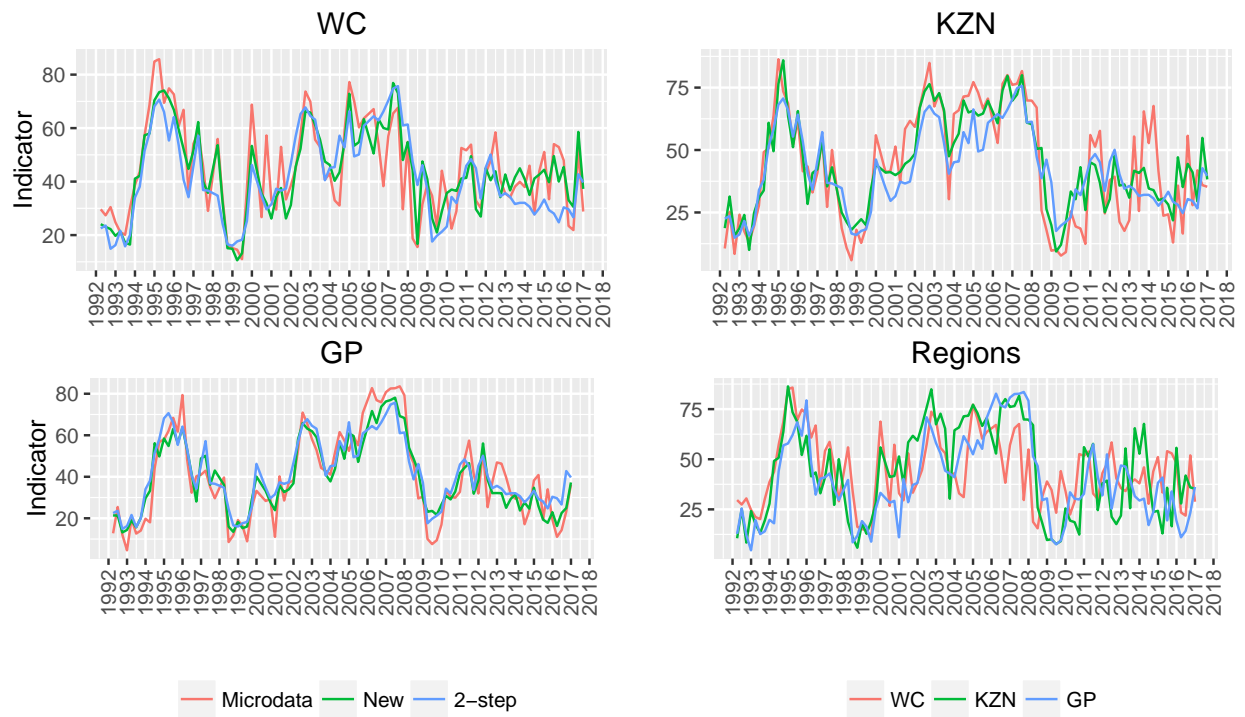


Figure 25: results

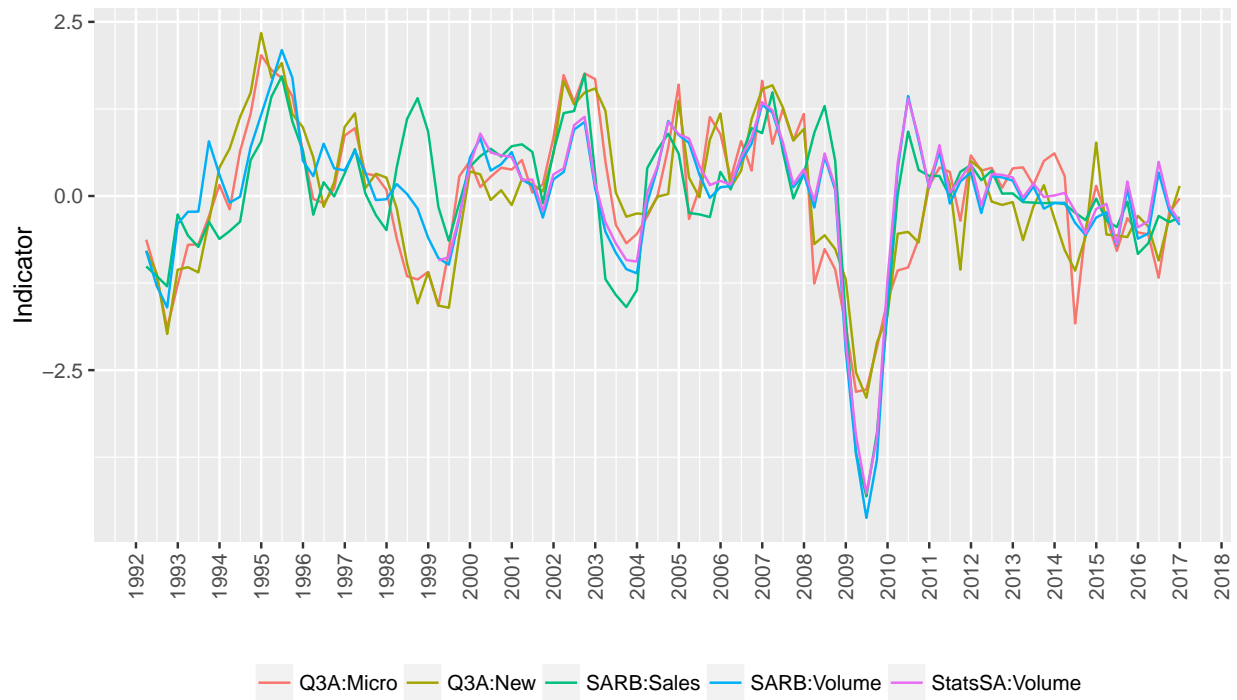


Figure 26: results

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

### 3.2.3 New Indicators

### 3.2.4 Constraints

## 3.3 Sectors

### 3.3.1 New BER classification: Q20 (Confidence)

### 3.3.2 Sectors2

### 3.3.3 Main industrial groups: Q20 (Confidence)

## 3.4 Regions

### 3.4.1 Regional Observations

### 3.4.2 Regional Results for Q20 (Confidence)

## 3.5 Reference Series

### 3.5.1 Total Manufacturing Activity

### 3.5.2 Correlations with reference series 26

### 3.5.3 Sectoral Manufacturing Activity

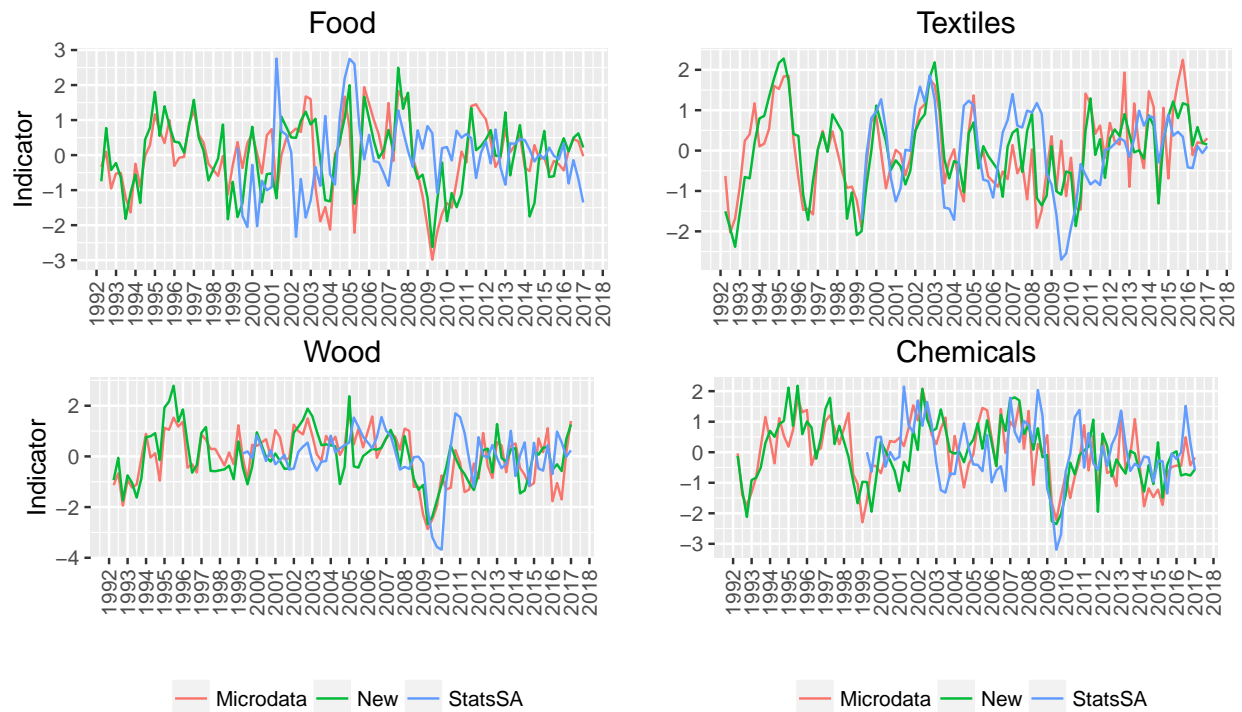


Figure 27: results

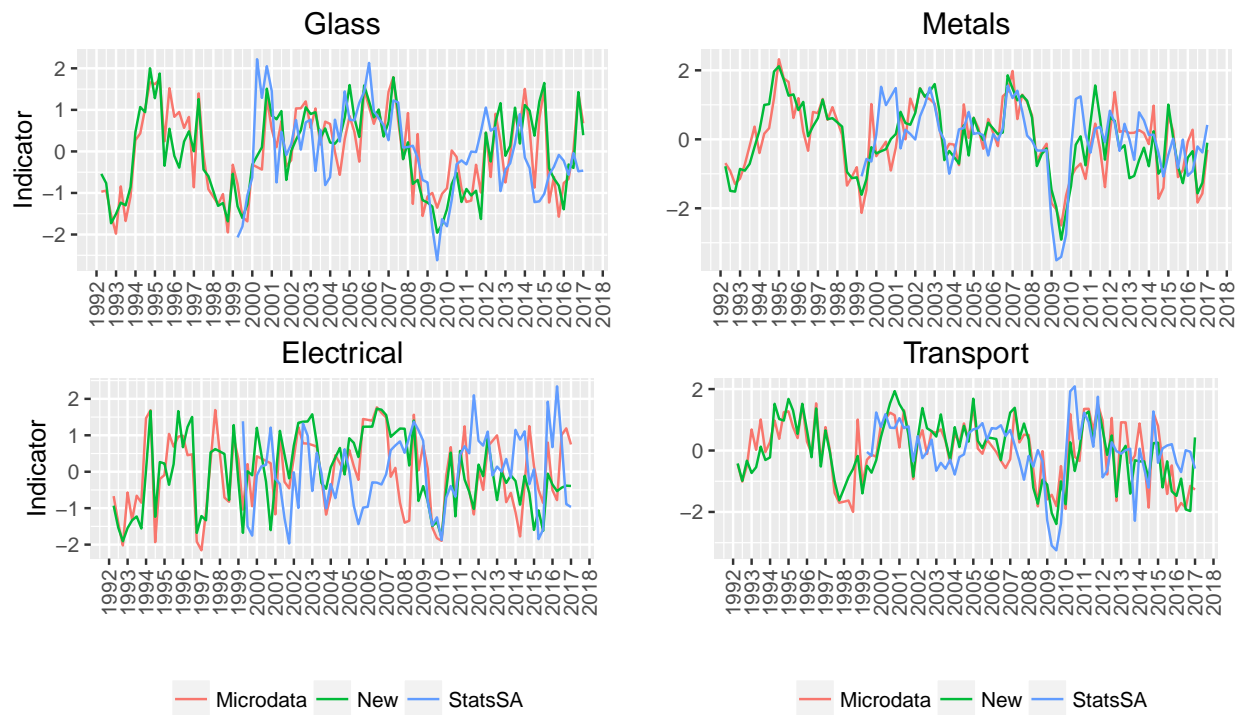


Figure 28: results

	Q1A	Q3A	Q1A_u	Q3A_u	Q1A_new	Q3A_new	Q1A_2s	Q3A_2s
Vol	27.18	24.98	22.99	21.58	23.20	21.71	24.45	22.66

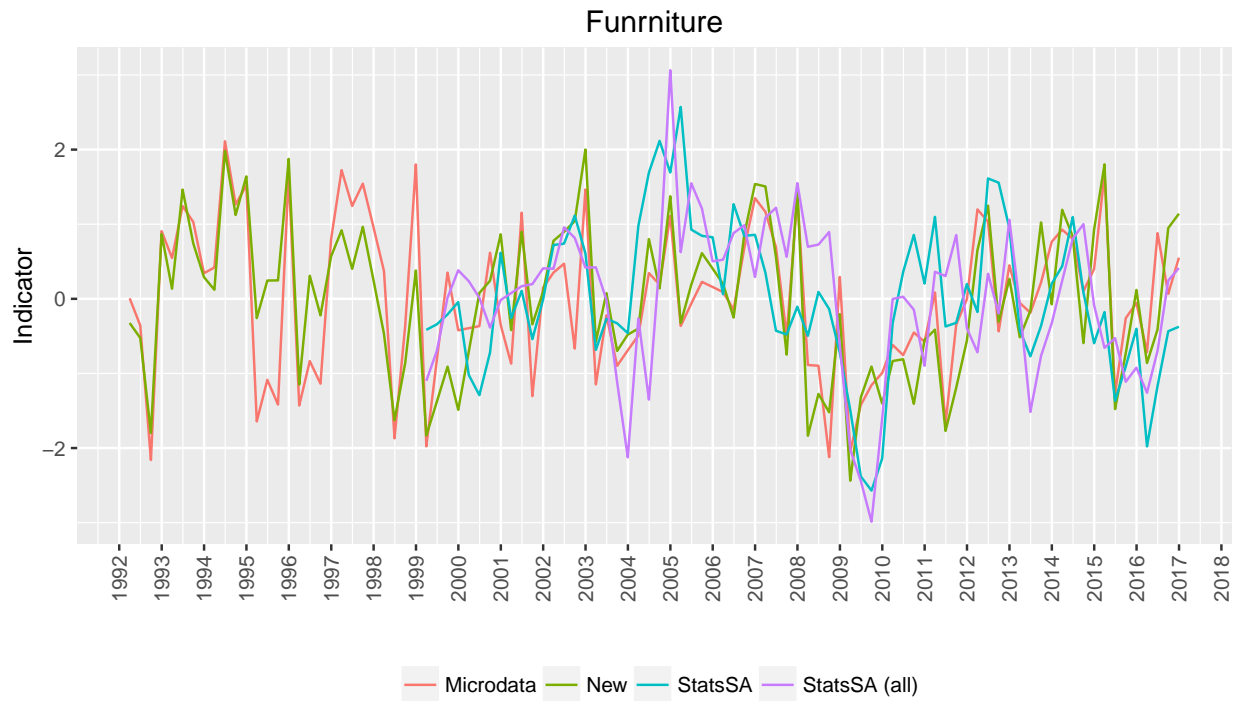


Figure 29: results

Ek het nou die mikrodata van 92Q2 tot 01Q1 wat jy uit die individuele files saamgestel het, by die historiese files (Retail\_92Q2-17Q2.xlsx en Wholesale\_19Q2-17Q2.xlsx) bygesit. Soos wat jy vir bou en vervaardiging gedoen het, moet jy eers toets of die resultate op grond van hierdie mikrodata en met die huidige faktore ooreenstem met die gepubliseerde data (op Quantec of sien BERdata\_Retail.xlsx en BERdata\_Wholesale.xlsx). Dit sal aantoon of daar êrens inkonsekwente sektor en faktor kodes ingesluit het.

In hierdie geval word die vergelyking egter deur die 1997Q1 en 2008Q3 hersienings van die opname data bemoeilik. In daardie kwartale het ons die sektor gewigte hersien. Om die ou en nuwe reekse aan mekaar te koppel, het ons ?? “adjustment” faktor by alle die kwartale voor die hersiening bygetel. (Die “adjustment” faktor is bereken as die absolute verskil tussen die resultate met die bestaande en nuwe faktors vir daardie kwartaal waarin die hersiening gedoen is en dieselfde waarde is net by die historiese reeks bygetel / afgetrek.) Hierdie faktors en die waardes van die reekse voor die hersiening is in worksheets gemerk met “revised” in BERdata\_Retail.xlsx en BERdata\_Wholesale.xlsx. Ek sou verwag dat jou berekenings vanaf die mikrodata gaan ooreenstem met die datareekse voor die hersiening.

Net die volgende sektore hoef bereken te word: . Wholesale: total, non-consumer goods (sektore 21xx) en consumer goods (22xx) . Retail: total, semi-durables (31xx), non-durables (32xx) en durables (33xx). . ?? Nuwe indeling moet bykom. Durables moet verdeel word tussen Hardware (3330) en Furniture, electronics & other (33xx excl. 3330). . Vanaf 1992 tot 2004 was daar ook ?? sektor 34xx. Hierdie sektor moet tot 2004 by die totaal bygetel word, maar hoef nie apart

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

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

gespesifiseer te word nie.

Sodra ek later kans kry, sal ek nuwe firma en sektor gewigte vir retail en wholesale bereken. Wanneer jy ook klaar daarmee is om enige sektor en faktor kode inkonsekwente reg te stel en kan aantoon dat die mikrodata met die gepubliseerde data ooreenstem, dan sal ek die sub-sektor kodes wat ons geskrap het (bv. 2140 en 2270 in die geval van die wholesale en 3170, 3340, 3350, 3360, 3410 en 3420 in die geval van die retail) te “convert” na hul ooreenstemmende nuwe sub-sektor kodes. Daarna kan jy die resultate herbereken met die nuwe faktore.

My berekening van die sektor gewigte sal ons help om lang verwysingsreekse (benchmarks) vir die sub-sektore te spesifiseer. Stats SA / SARB het lang tydreekse vir die totale, maar die ou retail sektor data is in 2003 gestaak en die nuwe data begin eers in 2005 in heersende pryse en vanaf 2008 in konstante pryse. Die wholesale sektor data is net in heersende prys vanaf 2005 beskikbaar. Stats SA se sektor indeling het ?? groot “other” komponent en die toedeling (al dan nie) daarvan tussen wholesale consumer and non-consumer goods aan die een kant en retail semi-durables, durables en non-durables aan die ander kant het ?? potensiele groot impak op die sub-sektorale groeikoerse.

Die pad vorentoe:

1. Jy moet bevestig dat die mikrodata met die gepubliseerde data ooreenstem. Waar daar verskille is, moet die sektor en provinsiale kodes in die file “Retail\_92Q2-17Q2” op Dropbox reggestel word.
2. Ek sal dan hierdie file gebruik om

Table 10: Retail Sector Codes

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

Table 11: Wholesale Sector Codes

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

- ou sektor kodes na die nuwes te verander (bv. 3150 en 3170 moet na 3160 verander word)
  - en die gewigte van 9 na 4 te verander (“ou” 1 tot 4 word mikro, 5 en 6 word small, 7 en 8 word medium en 9 word large)
  - die faktore vir die verskillende tydperke byvoeg.
3. Jy moet dan hierdie file met die nuwe gewigte gebruik om die reekse met mekaar en die verwysingsreekse te vergelyk.

Wat die verwysingsreekse (sien file: Retail) betref, daar is ?? breek in die reeks vir die komponente van 2004 tot 2007. Verder is dit nie nodig om “other” tussen die 3 hoofkomponente toe te deel nie, want dit het omtrent geen impak op die groei koers nie.

T1 Non-durables: Old 85-03: Food, groceries, beverages, pharmac & cosmetics & books. New 08+: gen dealers, spec food, pharmac & toiletries T2 Semi-durables: Old 85-03: Footw, men & women’s clothing, textiles, dom furnish, glass & sport. New 08+: clothing T3 Durables: Old 85-03: Furniture, appliances, audio, TV, jewellery & hardware. New 08+: Furniture & hardware

T3 Durables moet dan ook gedeel word tussen Hardware en Furniture & Other

## 4.2 The Wholesale Survey

## 4.3 The Motor Vehicle Survey

Ek het die mikrodata voor 2001 in een historiese file Motor\_92Q2-17Q2.xlsx bygesit. Sal jy asb. toets of die resultate op grond van hierdie mikrodata en met die huidige faktore ooreenstem met die gepubliseerde data (op Quantec of sien BERdata\_Motor.xlsx)? As dit ooreenstem, kan jy die impak van nuwe firma gewigte (wat ek nog moet bepaal) op die resultate bereken.

. Use NAAMSA total and passenger car sales as reference series

Ek het gekontroleer en die motor handel opname resultate is soos dié van die klein en groothandel in 96Q4 terug hersien. Jou berekenings vir 96Q4 - 98Q3 sal ook verskil van die gepubliseerde data. Sien die worksheet “Motor 86-96 revised” in die workbook “BERdata\_Motor” op Dropbox. Soos in die geval van die ander twee sektore, is die onderliggende mikro data dus korrek.

## 5 The Services Survey

Die mikrodata vir die dienste opname is beskikbaar vanaf die begin van die opname in 2005. Sal jy asb. toets of die resultate op grond van hierdie mikrodata ooreenstem met die berekende data (sien BERdata\_Services.xls; dit is nie op Quantec beskikbaar nie). Alhoewel ons ?? gewig vir elke medewerker het, het ons tot nou dieselfde (geen) gewig gebruik om die resultate van die sub-sektore te bereken. Die vraelyste en vraag kodes is in ServQues with codes.pdf As jy bevestig het dat die

Table 12: 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

mikrodata stem ooreen met die berekende data, dan sal ek die sub-sektor kodes wat ons geskrap het (bv. 6020, 6030, 8020, 8050, 8060, 8220, 8230, 8240 en 9060 ) “convert” na hul ooreenstemmende nuwe sub-sektor kodes. Ek moet nog die firma en sektor gewigte bereken. Ons kan daarna die impak daarvan en ?? totaal vir “ander dienste” bereken. Die werk oor die sektor gewigte sal my ook in staat stel om die verwysingsreeks te spesifiseer. Net die volgende sektore hoef (op die ou end) bereken te word: . Totaal (word tans weens ?? gebrek aan sektor gewigte nie bereken nie) . Hotels and restaurants (sektore 6xxx) . Transport and storage (7xxx) . Real estate (8000, 8010, 8020) . Business services (8xxx excl. dié van real estate)

Daar is te min respondente vir personal services (9xxx) om as ?? aparte sub-sektor te bereken. Ons sal dit waarskynlik ook nie in die totaal kan insluit nie.

Vir die periode 05Q2-08Q2 is “published sevice constraints” data nie deur 0.67 gedeel nie, d.w.s. die gepubliseerde data is uit 67 en nie uit 100 nie. Lyk my daar was een of ander berekeningsfout vir die periode 08Q3 - 09Q2 (?). Ek kan nie presies agterkom wat daar verkeerde is nie. Dit lyk amper of die berekening van net “transport en storage” en “business service” sektor na 100 verander is. Soos in bogenoemde geval, lyk dit of die mikro-data korrek is, al stem dit in hierdie geval nie met die “gepubliseerde” data ooreen nie. ?? Ander ding wat ek raakgeloop het en waarvan jy ?? aantekening in die meta dokument moet maak, is dat die woord “skilled” vanaf 15Q2 by die “shortage of labour constraint” vraag bygevoeg is. Die vraag was dus voor en na 15Q2 nie dieselfde nie en mens moet dit in ag neem as jy die tydreeks gebruik.

## 6 Conclusion

## References

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.