Alignment of the Quarterly Financial Statistics to the Annual Financial Statistics data

Pillay, Sagaren and de Beer, Joe

Statistics South Africa, Statistics South Africa

September 2016

Alignment of the Quarterly Financial Statistics to the Annual Financial Statistics data

Sagaren Pillay[1], Joe de Beer[2]

Abstract

Statistical data are often compiled at different frequencies. When analysing high and low frequency data on the same variable one often encounters consistency problems. In particular, the lack of consistency between quarterly and annual data makes it very difficult for time series analysis. This paper discusses the processes and challenges for the alignment of the quarterly and annual financial statistics surveys by industry. The process consists of three phases, the initial editing, to deal with large inconsistencies, a presentation of the methodology using the quarterly related series to interpolate the annual series, and an analysis of the results. In the initial editing phase the large differences are resolved by manually editing the input data and imputing for missing data. The temporal disaggregation/benchmarking technique used are based on the Fernandez optimisation method of allowing random drift in the error process. The main characteristic of this method is that quarter- to- quarter movements are preserved while quarterly-annual alignment is achieved. The diagnostics performed indicate that the Fernandez random walk model method produces plausible results.

Key Words: disaggregation, benchmarking, optimisation, Fernandez random walk model

1. Introduction

There is a great demand for consistent and coherent data by researchers, economists, and policy makers. In particular, for time series analysis high and low frequency data need to be plausible and meaningful. Given that the Quarterly Financial statistics (QFS) data is independent of the Annual financial statistic (AFS) data, differences between the annual estimates and the corresponding quarterly estimates occur naturally. Thus, the initial editing phase assists in flagging and resolving large differences and imputing for missing values. The small differences do not have a significant impact on the overall process and are processed mechanically. If the plausibility checks reveal that results are not satisfactory, they could be resolved by fine tuning the input.

The main characteristic of the benchmarking process is that quarter on quarter movements are preserved as much as possible, while enforcing annual alignment.

[1] Executive Manager: Private Sector Financial Statistics, Statistics South Africa, Pretoria, South Africa, Email: sagarenp@statssa.gov.za
[2] Deputy Director General: Economic Statistics, Statistics South Africa, Pretoria, South Africa, Email: joeDB@statssa.gov.za

There are mathematical and statistical methods that may be used for temporal disaggregation. The distinction between a mathematical and a statistical method is that a mathematical model treats the process of an unknown quarterly series as deterministic and treats the annual constraints as binding, whereas a statistical model treats the process of an unknown quarterly series as stochastic and allows the annual constraints to be either binding or not binding.

The most widely used mathematical benchmarking methods are the Denton adjustment method (Denton, 1971) whilst the most commonly used statistical methods are the Chow-Lin regression methods (1971) and their extensions (Fernandez, 1981; Litterman, 1983).The mathematical methods are easier to use, but they do not correct any serial correlation in the time-series data. The problem of serial correlation is a reality when dealing with economic time series data. The assumption of no serial correlation in the residuals of sub-annual estimates is generally not supported by empirical evidence (Chen, 2007).

The software package ECOTRIM, from Eurostats, was used for the temporal disaggregation.

2. The surveys

Each year in November, Statistics South Africa publishes the estimates of the Annual Financial Statistics (AFS) Survey. This survey collects a range of financial statistics in respect of enterprises in the formal business sector of the economy, excluding agriculture and hunting services, financial intermediation, insurance, government institutions and educational institutions.

The information is collected for the financial years of the enterprises that ended on any date between 1 July of a particular year and 30 June of the following year. The preliminary estimates are published with a lag of approximately twelve months and the previous year's estimates are revised each year. Thus results are made final after two years.

Discrepancies between successive AFS estimates may be attributed to numerous factors such as sampling design, changes in actual reporting, restructuring of large businesses, and exchange rates.

The Quarterly Financial Statistics (QFS) survey also collects financial statistics from enterprises in the formal business sector and excludes agriculture, hunting, forestry and fishing, financial intermediation and insurance and government institutions. The QFS estimates are published each quarter with a lag of approximately twelve weeks. Each quarter when the preliminary estimates are published, the previous quarter's estimates are revised. The revisions are mainly due to changes in reporting and late submissions of questionnaires.

When comparing the annual survey results and the related quarterly survey results it is important to keep in mind that the annual results are based primarily on information from audited financial statements whilst the quarterly results are based on estimates and proportional adjustments to financial statements.

The alignment of the AFS and QFS data is confined to the industries where the coverage and consistency of the data is comparable.

3. Methodology

The basis for the Fernandez method is provided by Chow and Lin (1971). The Chow-Lin method is a regression based method for interpolating, distributing, and extrapolating time series using related indicators. Let there be n observed annual observations y_1, y_2… y_n for each t=1, 2…, n we want to estimate quarterly values $x_{t,1}$, $x_{t,2}$, $x_{t,3}$, $x_{t,4}$. Assume that the series satisfies a linear stochastic relationship with a set of p observed quarterly variables. That is

$$x_{t,I} = z_{t,i}\,\beta_1 + z^2_{t,i}\,\beta_2 + \ldots + z^p_{t,i}\,\beta_p + u_t \qquad \textbf{(1)}$$

We can write the relationship between the estimated quarterly series and the observed related quarterly series as

$$x_t = z_t\beta + u_t\,, \qquad \textbf{(2)}$$

Where x is a Tx1 vector, z is a Txp matrix of p related series, β is px1 vector of coefficients, u_t is a Tx1 vector of random variables with mean zero and TxT covariance matrix V. Further, assume that there is no serial correlation in the residuals of the quarterly estimates.

The relationship between the annual and quarterly series, subject to the aggregation constraint, expressed in matrix form yields

$$y = B'x = B'\,(z\beta + u) = B'z\beta + B'u. \qquad \textbf{(3)}$$

Equation (3) is the temporal additivity constraint relating the quarterly series to the annual series.

The estimated coefficient $\hat{\beta}$ is the GLS estimator with y being the dependent variable and annual sums of the related quarterly series as the independent variables (Chow and Lin, 1971).

$$\hat{\beta} = [z'B(B'VB)^{-1}B'z]^{-1}z'B(B'VB)^{-1}y \qquad \textbf{(4)}$$

The linear unbiased estimator of x is

$$\hat{x} = z\hat{\beta} + VB\,(B'VB)^{-1}[y - B'z\hat{\beta}\,)]. \qquad \textbf{(5)}$$

The first term in (6) applies $\hat{\beta}$ to the observed related quarterly series of the explanatory variables. The second term is an estimate of the Tx1 vector u of residuals obtained by distributing the annual residuals $y - B'z\hat{\beta}$ with the TxM matrix VB $(B'VB)^{-1}$. This implies that if the quarterly residuals are serially uncorrelated, each with variance σ^2, then $V = \sigma^2 I_{TxT}$. The annual discrepancies are distributed in exactly the same fashion as Denton's basic model with $A = I_{TxT}$.

According to Litterman (1983), the Chow-Lin procedure with V proportional to the identity will be inadequate. This procedure may lead to step discontinuities of the quarterly estimates between years because it allocates each annual residual equally among the four quarterly estimates. Chow-Lin proposes a method to estimate the covariance matrix V under the assumption that the errors follow a first-order autoregressive AR (1) process. This, however, is inadequate only when the error process is stationary.

Fernandez (1981) proposed a generalisation of the Chow-Lin procedure based on a random walk model. According to Fernandez, the quarterly residuals follow the process

$u_t = u_{t-1} + \varepsilon_t$,

Where $\varepsilon_t \sim N(0, V_{TxT})$, is a vector of random variables with mean zero and covariance matrix V_{TxT}. From (3), the relationship between x and z can be written as

$x_t - x_{t-1} = z_t\beta - z_{t-1}\beta + u_t - u_{t-1}$ where , **(6)**

$x_t = z_t\beta + u_t$ and $x_{t-1} = z_{t-1}\beta + u_{t-1}$ **(7)**

Writing D as a difference operator, we have,

$Dx_t = Dz_t\beta + Du_t$, **(8)**

Given that the sum Dx_t is not equal to y, the relationship between the annual and quarterly series can be written in matrix form as

$\Delta y = QDx = QDz\beta + QDu$ **(9)**

Where Δy is MxM, and Q is MxT. This specification holds if the final sub-annual estimates x in year 0 are constant, an assumption considered reasonable for large sample size.

Given that $u_t = u_{t-1} + \varepsilon_t$ and setting $QD = B'$, we have

$\hat{x} = z\hat{\beta} + (D'D)^{-1} B (B' (D'D)^{-1} B)^{-1}[y - B'z\hat{\beta})]$ and,

$\hat{\beta} = [z'B(B'(D'D)^{-1}B)^{-1}B'z]^{-1}z'B(B'(D'D)^{-1}B)^{-1}y$

When $A = D'D$,

$\hat{x} = z\hat{\beta} + A^{-1} B (B' A^{-1} B)^{-1}[y - B'z\hat{\beta})]$ and,

$\hat{\beta} = [z'B(B'A^{-1}B)^{-1}B'z]^{-1}z'B(B'A^{-1}B)^{-1}y$

Are solutions to the first difference regression model.

In general, the Fernandez optimisation method requires:

1) Before estimating the sub-annual series through interpolation or distribution, the behaviour of the series should be studied. If the series is non-stationary and serially correlated, then the first difference data should be used to transform the data in order to obtain stationary and uncorrelated series;

2) If the first difference is not enough, other transformation is needed to convert residuals to serially uncorrelated and stationary variables; and

3) Given proper transformation, the degree of serial correlation can be tested by generalized least square estimation.

For this model, the parameter to be estimated is in the autocorrelation process of the errors. The ECOTRIM software option of having the parameter to be optimally determined in the estimation was used. This was done to ensure strong short-term movement preservation.

4. Estimation Results

The mining, manufacturing, construction, trade, transport, and community services industries were selected for analysis. For each of these industries the time series for turnover and capital expenditure were analysed.

The correlation coefficients between the QFS annual aggregates and the AFS estimates (Table 1) are between 0.96 and 0.99., and indication of strong correlation between the annual and the corresponding related series for each industry. This indicates that the related series does provide information on the short term movements for the estimated series.

Table 1 : Correlation coefficient ρ between AFS and aggregated QFS turnover

Industry	ρ
Mining	0.96
Manufacturing	0.96
Construction	0.98
Trade	0.99
Transport	0.99
Services	0.97

Table 2: Average absolute change in period-to-period Growth Rates

Industry	*Turnover*	*Capital Expenditure*
Mining	0.048421	0.161809
Manufacturing	0.038397	0.134497
Construction	0.016199	0.121970
Trade	0.012586	0.133587
Transport	0.012805	0.055663
Services	0.037797	0.904999

The short-term movement preservation can be measured in terms of level, proportion, and growth rates. For each of the estimated series the average absolute change in period-period growth rates were computed. The statistic was computed as follows:

$$c = \sum_{t=2}^{T} \left| [(x_t / x_{t-1})/(z_t / z_{t-1})] \right| / (T-1)$$

Where x_t is the estimated series, and zt the related series. This statistic measures the changes in the period- to- period growth. The average absolute change in period- to-period growths (Table 2) is much larger for capital expenditure. This can be observed in the graphs for the related and estimated series for all of the industries (Tables 2, 4, 6, 8, 10, and 12).

For all the series shown below the estimated QFS quarterly estimates aggregates to the AFS series. The final estimates for turnover are fairly close in level to the related series whilst the final estimates for capital expenditure displays more divergence from the related series for all the estimates. This may be due to large differences in capital expenditure estimates between the quarterly aggregates and the annual series. The differences between the AFS and QFS totals for capital expenditure estimates may be attributed to poor reporting of quarterly capital expenditure by large businesses on one hand and the volatility in the estimates for the smaller sampled units.

Table 3: Grouped average absolute change at break points

Industry	*Turnover at breaks*		*Capex at breaks*	
Mining	0.047943	0.058307	0.139756	0.170621
Manufacturing	0.023039	0.063252	0.081363	0.214963
Construction	0.011144	0.025912	0.18851	0.152179
Trade	0.012325	0.014919	0.129848	0.099714
Transport	0.013334	0.014302	0.103382	0.056989
Services	0.050119	0.037294	0.919541	0.887717

A good method for temporal disaggregation should not generate final estimates that impose large distortion at the beginning and ending periods of the sample. Such distortion is measured by inconsistent period-to-period growth rate between the final estimates and the related series (Chen, 2007). The two grouped averages were calculated for the periods before and after the samples were introduced. The final estimates for manufacturing and transport capital expenditure exhibit some degree of distortion at breaks between years (Table 3).

5. Limitations

The definition of turnover is not standard between the two surveys. Turnover in the QFS only includes sales revenue and income from services rendered to all the industries except SIC 8 for which rental income on land and buildings, rental income from plant and equipment and interest received are included in turnover. The AFS turnover includes sales revenue, income from services rendered, income from mineral rights leases, income from rental and leases of land, buildings and other structures under operating leases, income from leasing and hiring plant, machinery and equipment, and income from leasing and hiring of motor vehicles and other transport equipment.

Different sampling frames and methodology are used for comparative periods of AFS and QFS. The sampling specification according to SIC digit level for he different industries differs for the two surveys. The QFS estimates for SIC 8 include SIC 8 include SIC 8899 while estimates of the AFS do not include this sub-sector.

The cut-off points for size group differ significantly between AFS and QFS due to different sampling methodologies, different cut-off points results in different size group allocation for certain enterprises.

The reference periods differ. The QFS collects data based on calendar quarters while the AFS collects based on financial year end

Classification differences occurred between the two sampling frames for AFS and QFS, which contribute to the differences in estimates.

Due to the QFS being a much smaller sample than AFS, different cut-off points for corresponding size groups were applied which resulted in differences in weights.

Inter divisional transfers within enterprises may be present in the QFS reported data while for the AFS such values are eliminated and resulted in lower values.

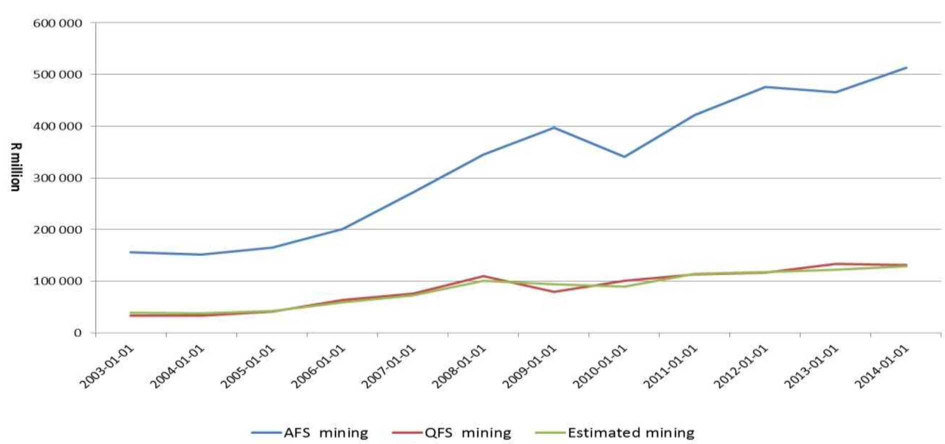


Figure 1 : Quarterly, annual and estimated turnover series for mining

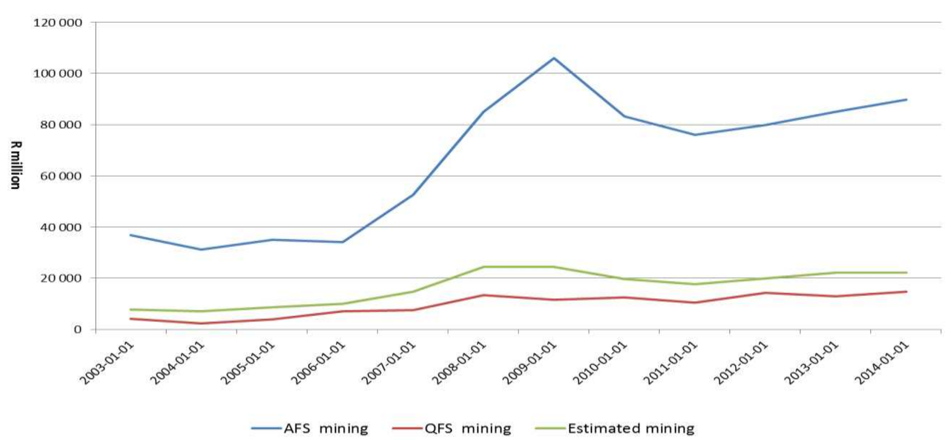


Figure 2 : Quarterly, annual and estimated capital expenditure series for mining

Figure 3 : Quarterly, annual and estimated turnover series for manufacturing

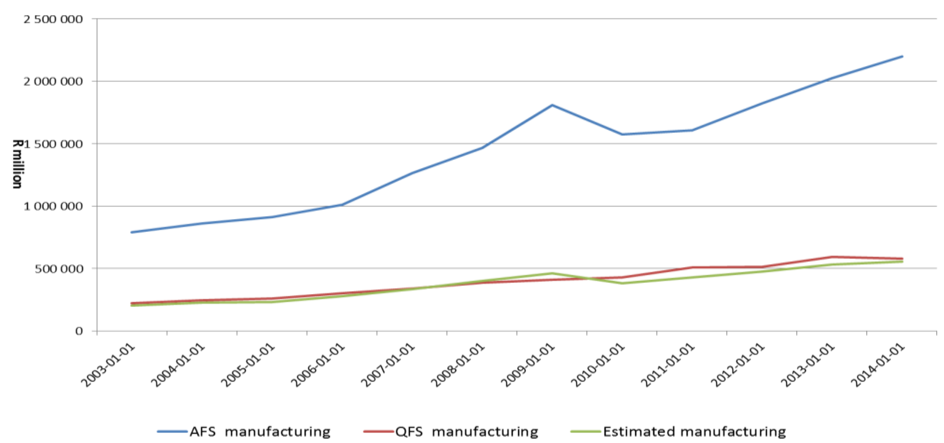


Figure 4 : Quarterly, annual and estimated capital expenditure series for manufacturing

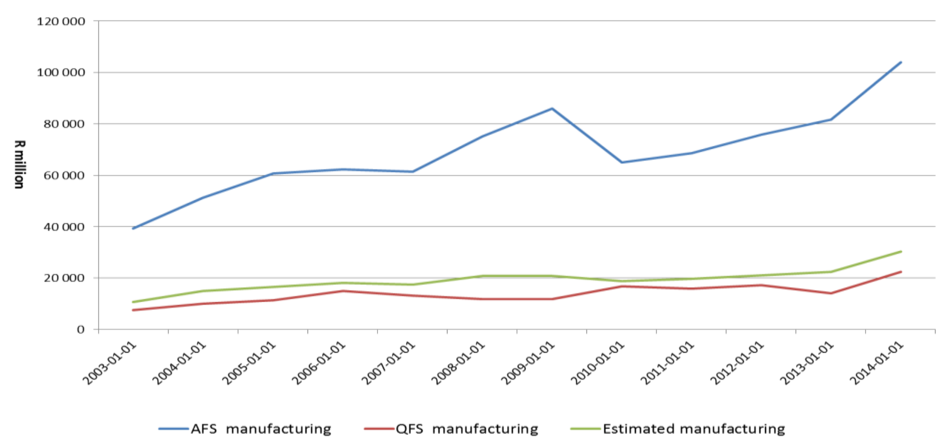
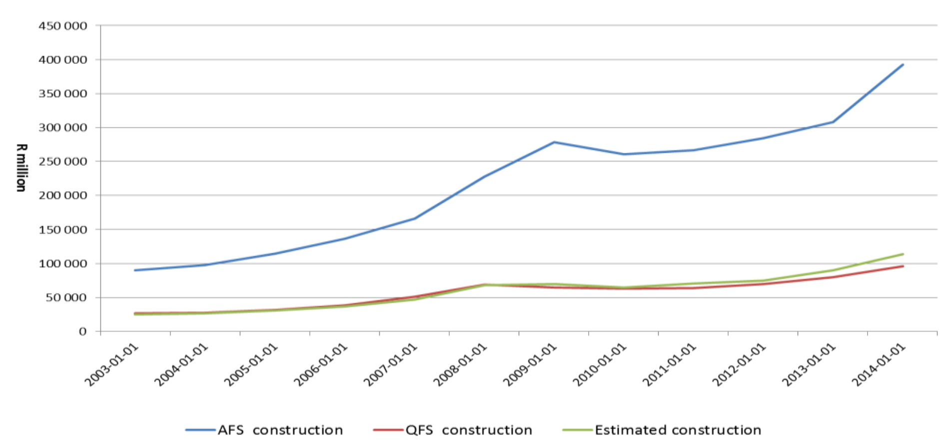


Figure 5 : Quarterly, annual and estimated turnover series for construction



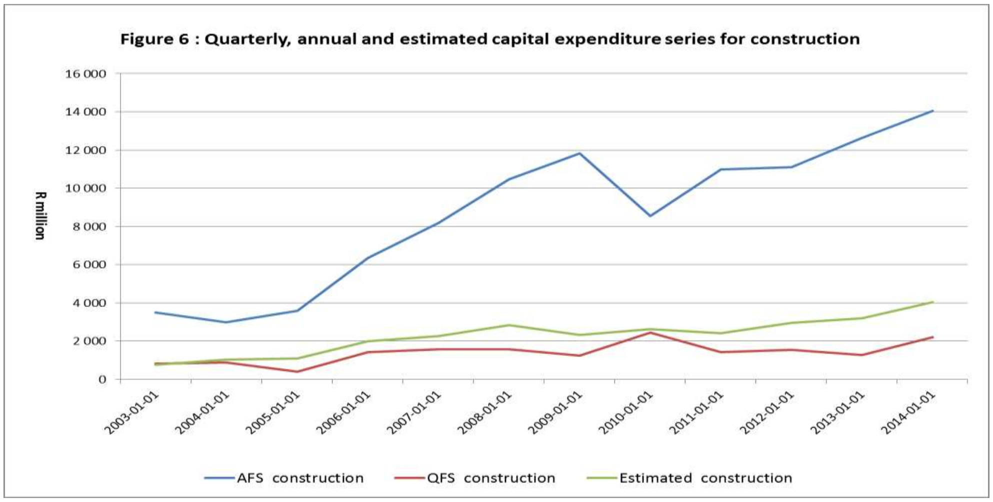


Figure 6 : Quarterly, annual and estimated capital expenditure series for construction
16 000
14 000
12 000
10 000
8 000
6 000
4 000
2 000
0
R million
2003-01-01
2004-01-01
2005-01-01
2006-01-01
2007-01-01
2008-01-01
2009-01-01
2010-01-01
2011-01-01
2012-01-01
2013-01-01
2014-01-01
AFS construction
QFS construction
Estimated construction

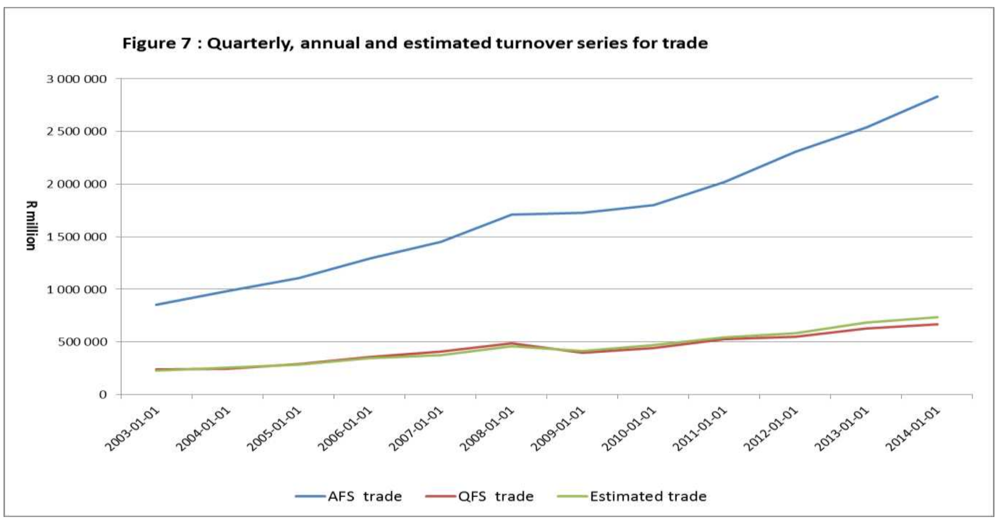


Figure 7 : Quarterly, annual and estimated turnover series for trade
3 000 000
2 500 000
2 000 000
1 500 000
1 000 000
500 000
0
R million
2003-01-01
2004-01-01
2005-01-01
2006-01-01
2007-01-01
2008-01-01
2009-01-01
2010-01-01
2011-01-01
2012-01-01
2013-01-01
2014-01-01
AFS trade
QFS trade
Estimated trade

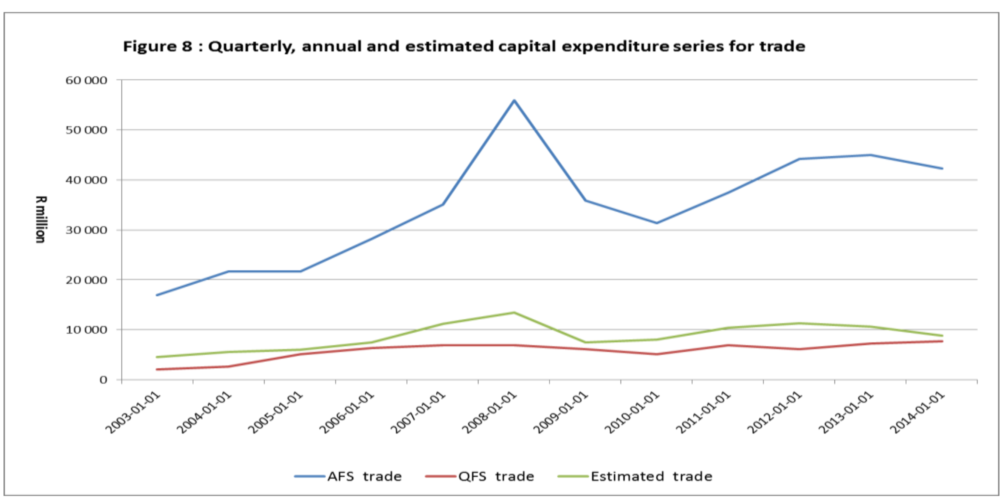


Figure 8 : Quarterly, annual and estimated capital expenditure series for trade
60 000
50 000
40 000
30 000
20 000
10 000
0
R million
2003-01-01
2004-01-01
2005-01-01
2006-01-01
2007-01-01
2008-01-01
2009-01-01
2010-01-01
2011-01-01
2012-01-01
2013-01-01
2014-01-01
AFS trade
QFS trade
Estimated trade

Figure 9 : Quarterly, annual and estimated turnover series for transport

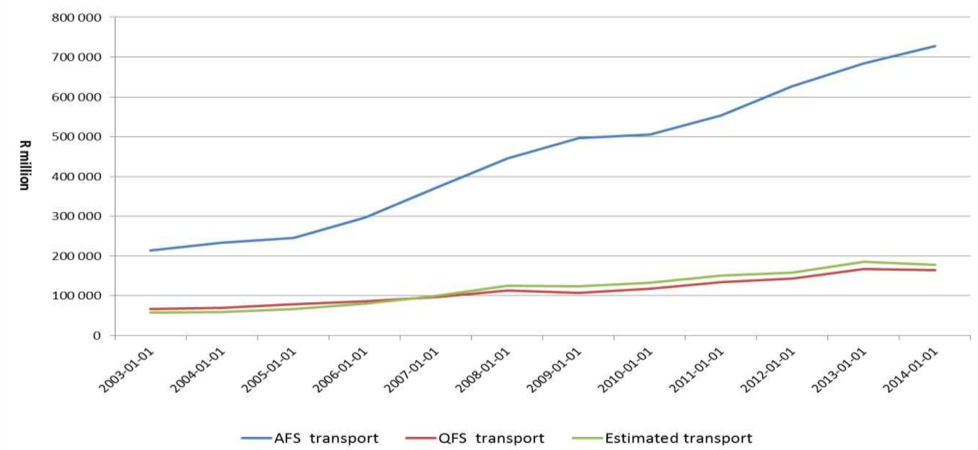


Figure 10 : Quarterly, annual and estimated capital expenditure series for transport

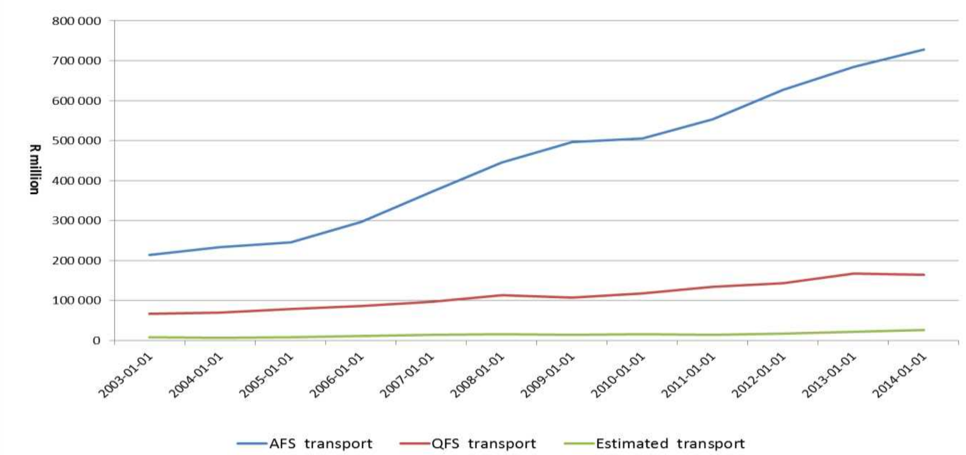


Figure 11 : Quarterly, annual and estimated turnover series for community, social and personal services

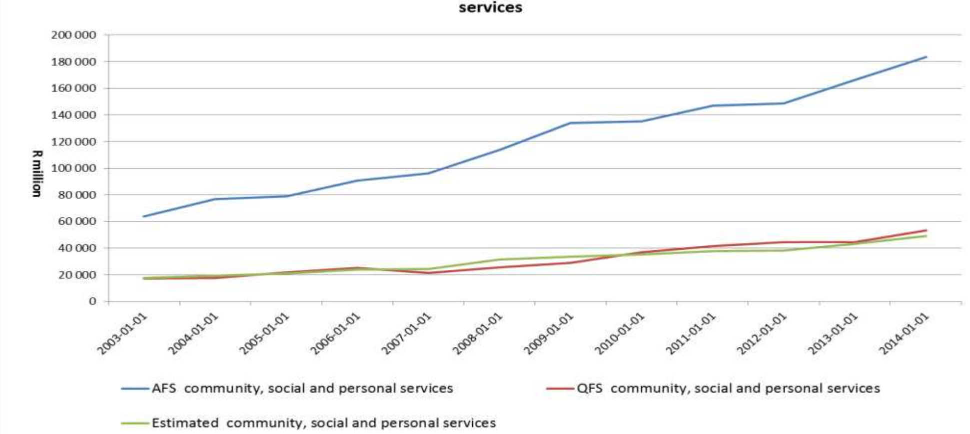
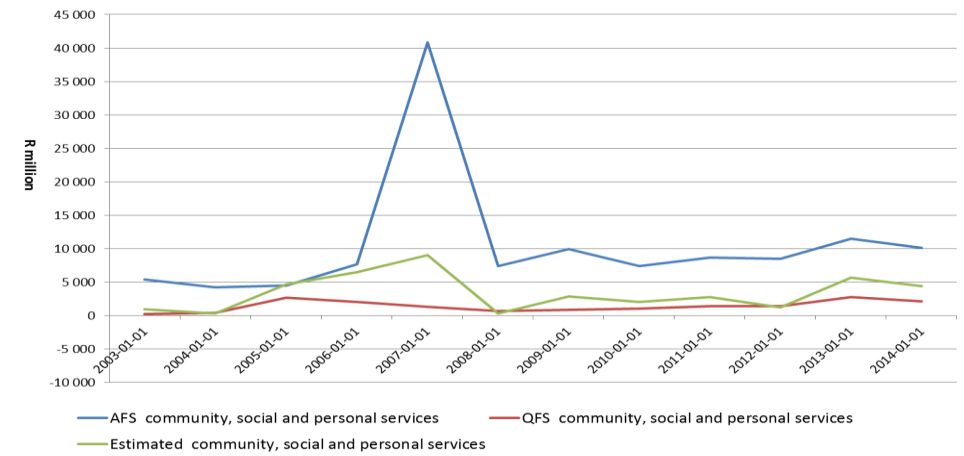


Figure 12 : Quarterly, annual and estimated capital expenditure series for community, social and personal services



6. Conclusion

In our attempt to align the AFS and QFS data we have found that the benchmarking exercise for the turnover variable has produced plausible results notwithstanding the several limitations that impact on this process. This is true for all the industries investigated. With respect to the capital expenditure series, the volatility and large differences between the QFS aggregates and AFS estimates has led to a poor estimation results for all of the industries (Figures, 2, 4, 6, 8, 10, 12).

The current QFS survey needs to focus on improving the quality of the estimates for capital expenditure to the extent that there is better comparability with the AFS capital expenditure estimates.

References

Denton, F.T., (1971), "Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization," *Journal of the American Statistical Association* 66(333), pp. 99-102.

Chen, B., (2007), "An Empirical Comparison for Temporal Distribution and Interpolation at National Accounts," Bureau of Economic Analysis, Washington, DC

Chow, G.C., Lin, A.L. (1971), "Best linear unbiased interpolation, distribution and extrapolation of time series by related series," *The review of Economics and Statis*tics 53, pp. 372-375.

Eurostat (1999), Handbook on Quarterly National Accounts, Office for Official Publications of the European Communities.

Fernadez, R.B. (1981), "A methodological note on the estimation of time series," *The Review of Economics and Statistics* 63(3), pp. 471-476.

Litterman, R.B. (1983), "A random walk, Markov model for the distribution of time series," *Journal of Business and Statistics*, 1, pp. 169-173.