Time series for main variables on the performance of Dutch SMEs

Methodology description

Ton Kwaak
Werner Liebregts

Zoetermeer, February 2012

EIM onderdeel van Panteia

SCALES
SCientific AnaLysis of Entrepreneurship and SMEs
www.entrepreneurship-sme.eu

This research has been financed by SCALES, SCientific Analysis of Entrepreneurship and SMEs (www.entrepreneurship-sme.eu)

EIM Research Reports	
Reference number	
Publication	February 2012
Number of pages	28
E-mail address corresponding author	akw@eim.nl
Address	EIM Bredewater 26 P.O. box 7001 2701 AA Zoetermeer The Netherlands Phone: +31(0)79 322 22 00

All the EIM research reports are available on the website www.entrepreneurship-sme.eu.

Contents

Summary

National Accounts provide detailed information on the development of the economy detailed by sector of industry. However, a disaggregation by enterprise size class is not available. The available data on the size class structure of the economy shows huge and unrealistic fluctuations and can therefore not directly be used to disaggregate the information from National Accounts. This paper reviews some methods to smooth developments shown in source data with respect to the share of small, medium-sized and large enterprises. It appears that in principle, the Hodrick-Prescott filter is most suited to fulfill this task.

A modified Hodrick-Prescott filter is used. In particular, the simultaneous smoothing of interrelated series, taking into account the definitional relations existing between them, performs quite well.

The methodology has been tested for two rather different sectors of industry, *i.e.* chemical industry (large-scaled, business-to-business oriented, capital-intensive) and retail trade (small-scaled, oriented towards consumers, labor-intensive). It appears that adjusted series, according to industry experts, give a more realistic description of historical developments than the original series do.

1 Introduction

An analysis of the development of the size class structure of the development of the business economy requires time series on various items, such as sales, value added, employment and investment, disaggregated by sector of industry and enterprise size class. Regarding the size class structure of the business economy, such data are annually reported in the so-called Production Statistics (in short: PS) by Statistics Netherlands[1]. However, these data appear to be less suitable for a year-on-year analysis. A year-on-year analysis is feasible using data from the National Accounts (NR)[2]; this data source lacks a distinction by enterprise size class though[3]. Hence, in order to carry out an analysis regarding the development of SMEs and large enterprises over time, the PS data should be combined with the data coming from the NR. The most straightforward method to do this is to apply the distribution by enterprise size class from PS data to the NR data. It appears that this method leads to many annual fluctuations on the size class level that are deemed implausible according to sector experts. Sector experts stress that these fluctuations involve statistical breaks in the time series instead of actual developments[4]. Apparently, the PS data by sector and size class needs further processing to filter out such breaks. A method has been developed to extract trends from noise in the PS data. After removing the noise, the adjusted PS-data can be used to distribute NR data by sector of industry across enterprise size classes. This report explains how this is done. The final result of applying this methodology has published[5].

The report proceeds as follows. First, chapter 0 argues that from the most common filter (smoothing) methods, the Hodrick Prescott (HP) filter is the most appropriate one. Then, chapter 3 elaborates on this method adapting it to the specific nature of the data under consideration. Chapter 4 describes the dataset to which the method is applied in terms of its classifications. To illustrate the functioning of the method in practice, chapter 5 shows some results for two rather different sectors of industry, *i.e.* the chemical industry and the retail trade sector.

[1] These data are closely related to the Structural Business Statistics published by EUROSTAT.

[2] Also composed by Statistics Netherlands.

[3] However, see (in Dutch) Pommée and Van Dalen (1997).

[4] For instance, enterprises that move from one size class to another may be assigned to the new size class only with a lag. Furthermore, the way smaller enterprises are observed differs from the method of observation of larger enterprises. Finally, changes in the survey's questionnaires have occurred over time.

[5] Main SME data (In Dutch: Reeksen kerngegevens MKB (incl. prognoses)), developed in the framework of the long term research program on small and medium-sized enterprises (SMEs) and entrepreneurship, which is being financed by the Dutch Ministry of Economic Affairs, Agriculture and Innovation; see http://data.ondernemerschap.nl/webintegraal/userif.aspx?SelectDataset=14&SelectSubset=41&Country=NL

2 Smoothing methods

2.1 Overview of methods

There are many methods smoothing time series; all methods involve some sort of averaging. This ensures that individual observations that can be considered as outliers (*i.e*. having measurement errors) cancel out against one another. Baxter and King (1999) have listed some objectives that a smoothing method should ideally fulfill. Apart from being operational, two are relevant here. First, the filter should leave the trend characteristics of the series unchanged. Second, the filter should preserve the turning points of the original time series, *i.e*. a phase shift should not occur.

In this chapter, the most common smoothing methods and their advantages and disadvantages are discussed. It should be stressed that this is by no means an exhaustive enumeration, but simply an enumeration of the most widely used smoothing methods. None of the methods perfectly fulfill the above-mentioned objectives and as a consequence none of them is free from controversy. Their popularity however, is not necessarily due to the fact that they are considered to be the optimal method (should one exist), but rather because of their suitability to a specific problem. It is up to researchers to decide upon which criteria are the most relevant in a certain application.

One problem encountered with all smoothing methods is the so-called end-point problem. In some way, all smoothing methods calculate a moving average of the original series. So, strictly speaking smoothing methods can produce for all but the few first and last observations. Analysts practically solve this problem by extrapolating the time series by some appropriate method.

Five types of smoothing methods can be distinguished: direct moving average methods, the Baxter-King filter, the Henderson filter, the Savitzky-Golay filter, and the Hodrick-Prescott filter.

2.1.1 Direct moving average smoothing

Simple (or unweighted) moving average smoothing is probably the most common method of smoothing time series, mainly due to its ease of implementation. However, it faces several major disadvantages. One of them is the fact that a moving average always lags the last observation, such that the moving average can be disproportionally affected by old observations dropping out of the average. The end-point problem applies.

Besides these two disadvantages, (arithmetically) weighted moving average smoothing encounters another disadvantage. Namely, the results are strongly affected by the choice for the weighting factors. Although the best possible weighting factors should be determined iteratively when using this method, the weighting factors are usually chosen in such a way that more weight is given to more recent observations. In general however, the choice for the weighting factors is relatively arbitrary.

Whereas weighted moving average smoothing assigns arithmetically increasing weights over time, exponential smoothing requires weighting factors to decrease exponentially as observations are getting older. That is, exponential smoothing also gives more weight to more recent observations. Exponential smoothing is also generally accepted as a useful tool for forecasting and it appears in single, double and triple form. Single exponential smoothing is easily applied, but not useful when there appears to be a trend in the corresponding time series. Then, one often applies double exponential smoothing, but regarding this method it is recommended to avoid its use when a seasonal trend occurs. Triple exponential smoothing (sometimes referred to as Holt-Winters smoothing) can be seen as the most extensive form of exponential smoothing, because it takes into account both seasonality and trends.

Also for exponential smoothing it holds that the choice for the smoothing factors is rather arbitrary. It is reasonable though to pick the value that minimizes de mean of squared errors (MSE). Exponential smoothing is often used as a forecasting technique. Although it also generates a trend for the observed time series, it is less common to use it for this purpose specifically.

A related alternative is median smoothing, which uses medians instead of means. An advantage of this technique is that the results are less biased by outliers within the smoothing window as compared to moving average smoothing. Hence, on the one hand median smoothing produces more reliable trend lines when outliers occur. On the other hand, median smoothing does not allow for assigning weighting factors. Autoregressive (integrated) moving average models are related to time series analysis as well, but they mainly serve as a way to better understand the data and to predict future values in the time series (forecasting). Therefore, these models are out of the scope of this overview of smoothing methods.

2.1.2 Baxter-King filter

The Baxter-King (BK) filter is a type of band-pass filter extracting the business cycle component of macroeconomic time series. Whether this component represents a reliable estimate of the real business cycle is still up to debate, particularly because there is no clear definition of the business cycle. Baxter and King define this cyclical component to be fluctuations lasting no fewer than six quarters and no more than thirty-two quarters (adopted from the definition of the business cycle by Burns and Mitchell, 1946). With their filter Baxter and King aim at eliminating trends and irregular components, while retaining business cycle components. Baxter and King recommend a twenty-four quarter moving average, consisting of three years of past data and three years of future data as well as the current year, when applying the BK filter to both quarterly and annual data. It uses symmetric weights to lags and leads, which makes sure that no phase shift occurs in the filtered series.

The fact that the BK filter can create spurious cycles at business cycle frequencies is often seen as an important disadvantage. It violates the first objective, as proposed by the authors themselves (*"the filter should leave the characteristics of the business cycle component unchanged"*, see section 2.1). A solution might be to increase the filter length, but one often has to deal with time series of limited length. A more appropriate solution could be the replacement of the truncated filter weights by some modified weights. Woitek (1998) argues that the resulting modified BK filter is preferable under specific circumstances. For the sake of completeness it should be mentioned that the BK filter also suffers from the endpoint problem, which limits the applicability even more when dealing with time series of limited length.

2.1.3 Henderson filter

The Henderson filter, named after Robert Henderson, can generate trend estimates out of observed time series. In short, to apply the filter one should minimize the sum of squares of the third difference of the moving average series. The Henderson filter is especially suitable for economic time series[6]. In obtaining the weights of the Henderson filter, as with all filters discussed in this section, a compromise is struck between retention of the curvatures of the original time series and sufficient smoothing. One of the advantages of the Henderson filter is that it does not affect cyclical structures appearing in the observed time series. Moreover, the filter eliminates almost all the irregular variations occurring, with a maximum length of six months.

The Henderson filter seems to be useless in case of annual data. The length of the Henderson filter depends on the variability of the data, and typically, a five or seven term filter is used for quarterly data. Then, a smaller filter length should be applied to annual data, but this will lead to insufficient smoothing of the original time series.

2.1.4 Savitzky-Golay filter

The Savitzky-Golay (SG) filter, named after its inventors, is a particular type of a low-pass filter and carries out a local polynomial regression on the point to be smoothed as well as several neighboring points. In the case of moving averages a least-squares fit is made to a zero order polynomial (*i.e.* a straight horizontal line or a constant value), whereas a SG filter performs a least-squares fit to a higher-order polynomial. The polynomial order should be higher, the wider the smoothing window. If the data in a particular smoothing window fits to a parabola, then it is preferred to use a second-order (quadratic) SG filter. The use of a fourth-order (quartic) SG filter is quite common in case the data fits to a fourth-order polynomial.

The major advantage of this method is the preservation of important features of the original time series, like the relative widths and heights. Usually, these features are flattened by other (simpler) averaging techniques.

If the data is irregularly spaced, a least-squares fit should be done within a moving window around each data point. This is computationally burdensome, particularly when the number of data points to the left and right is large, leading to the use of a higher-order polynomial. As an alternative, the user of the SG filter may pretend that the data points are equally spaced.

2.1.5 Hodrick-Prescott filter

The HP filter is said to be the most widely applied method to extract a trend component from an observed macroeconomic time series. It is a specific form of the Kalman filter, which is a useful algorithm in finding the optimal estimate of an unobserved component. The objective function contains a smoothing parameter lambda for which it holds that the higher its value, the smoother the trend

[6] If applicable, the original series should be seasonally adjusted.

estimate will be. Kalman filters have the useful property that they can be modeled such that they achieve certain objectives. For example, dummy variables can be added to the measurement equation, to restrain the effect of outliers or breaks on the trend estimate. Although other filters (*e.g.* the Henderson filter) also allow a certain degree of flexibility, the HP filter can be seen as a more comprehensive tool to fit a trend to time series data.

However, the HP filter still faces some disadvantages. First of all, like the BK filter, the HP filter is likely to generate spurious cycles and HP filtered series might show spurious correlation. The HP filter can be modified in a way which largely overcomes this effect. Second, the occurrence of a permanent shock in the observed time series induces a shift in the smoothed time series, which does not really exist. However, this effect can be undone by temporarily eliminating the acceleration term. When these kinds of shocks do not occur at all, the HP filter in its original form is still suitable. Third, the choice for the smoothing parameter is quite arbitrary and in general only based on previously chosen values in the existing literature. Although Hodrick and Prescott (1997) originally suggested and substantiated values for the smoothing parameter (100 for annual, 1.600 for quarterly and 14.400 for monthly data), endogenous determination of the smoothing parameter by others leads to rather different values. For example, Harvey and Trimbur (2008) argue that the smoothing parameter is often chosen too small. Indeed, Ravn and Uhlig (2002) recommend using a value of 129.600 for monthly data. On the other hand, for annual data, a value between 6 and 7 is suggested by Ravn and Uhlig (2002). Fourth, the noise in the data should be white noise, *i.e.* normally and independently distributed. Often this is assumed to be true, but in practice this may be doubtful. In that case, an irregular component will be subsumed to the cyclical component.

2.2 Most appropriate method

In the present application, a method that removes irregular changes in the time series is required. The method should be fast, flexible and easy to implement in a multi-variable context (in particular see section 3.2). Although Baxter and King (1999) admit that the HP filter is a reasonable approximation of the ideal business cycle filter, they argue that their own-developed high-pass filter is preferable when using quarterly data due to its greater ease of application.

In the current context, the HP filter is considered the most appropriate smoothing method. A review of the literature shows that the HP filter is the most common and widely applied method to detrend macroeconomic time series. Chuan (2006) even notes that the HP filter has almost become the default filter. Also, it has "withstood the test of time and the fire of discussion remarkably well" (Ravn and Uhlig, 2002:371). Ravn and Uhlig (2002) provide a summary of all critics and conclude that "none of the shortcomings and undesirable properties is particularly compelling" (also see Ahumada and Garegnani, 2000). Results should be interpreted carefully though, especially because the HP filter shows some unusual behavior of cyclical components at both ends of a time series. Therefore, it is recommended to drop at least three data points at both ends of each time series when using the HP filter on annual data. Regarding this it is noteworthy that the HP filter then drops as many observations as the BK filter. Alternatively, if available, forecasts could be added to the series. Another practical problem that users of the HP filter encounter is what value to choose for the smoothing parameter, but that is simply a matter of trial and error. Apart from a nonnegativity constraint, the smoothing parameter lambda can take any value. Thus, in practice, one is free to set lambda such that it fits the specific situation best.

The filter methods discussed so far have actually been developed for single time-series. In the current application, several interrelated series have to be processed[7]. This requires adaptation of the original filter method to the current application, Therefore, such adaptation being possible is an additional requirement on the method to be chosen. As shown in chapter 3, the HP filter suits these specific requirements and hence, the HP filter is also considered to be the most appropriate method for this particular problem of estimating the development of the size class pattern of the business sector, as is the purpose of this paper.

[7] Variables are connected to one another *via* definitional relations. Furthermore, as the *distribution* of phenomena across enterprise size classes is the main concern of this study, the sum of observations across size classes is given (at 100%) for each variable.

3 Model used

3.1 Formal representation of the HP-model

The HP filter is looking for the smoothest possible line through a number of data points. If $x_{i,t}$ denotes the original value of variable i in year t and if $y_{i,t}$ denotes the value for the adjusted series, then the y-series is generated by solving the following optimization problem:

$$\min_{y_{i,t}} \quad D_i = \Sigma_T (x_{i,t} - y_{i,t})^2 + \lambda \cdot \Sigma_T \{(y_{i,t+1} - y_{i,t}) - (y_{i,t} - y_{i,t-1})\}^2$$

The objective function above consists of two parts. First, the term $\Sigma_T(x_{i,t} - y_{i,t})^2$ describes the goodness of fit of the adjusted series to the original data series[8]. First, the irregular component $c_{i,t}$ that should be removed from the series is defined as $x_{i,t} - y_{i,t}$, such that the first term could also have been $\Sigma_T c_{i,t}^2$. Second, the term $\Sigma_T((y_{i,t} - y_{i,t-1}) - (y_{i,t-1} - y_{i,t-2}))^2$ is a measure of smoothness of the development as indicated by the y-series. This is because $y_{i,t} - y_{i,t-1}$ is the change of the y-series between time t-1 and time t, and similarly, $y_{i,t+1} - y_{i,t}$ is the change of the y-series between time t and t+1. Thus, the second part of the objective function will make the difference between these changes as small as possible. The parameter λ denotes the relative weight assigned to smoothness as compared to the fit to the original data series, and is restricted to be nonnegative. The adjusted series y approximate the original series if λ approaches zero, while the trend becomes a straight line if λ tends to infinity. Like Ravn and Uhlig (2002) suggest, a value of $\lambda = 7$ is used here.

3.2 Adaptation to the current problem

The HP filter has been designed for individual series. The current problem however, is about time series of elementary variables (like sales of four distinguished categories) and time series of composite variables, which are derived from these elementary variables (like total sales, being the sum of the sales of the four categories distinguished). As a result, not every D_i will be minimized individually (only elementary variables), but the sum of the D_i for all variables (both elementary and composite variables) by choosing the values of the y-series only for the elementary variables (see chapter 4 for variables).

Since the PS data are only used to distribute the sector totals of the NR data across three enterprise size classes (small, medium and large enterprises), the series x are defined in terms of shares of the individual size classes in the sector totals. Then, the sum of the adjusted y-variables across size classes should equal 100, so with three size classes in total, there are only two independent size classes.

[8] In a regression equation $x_t - y_t$ would be the residual of the equation. Besides, the least squares estimator just minimizes de sum of squared residuals $\Sigma_T(x_t - y_t)^2$.

Note that this procedure neglects the first part of the objective function regarding the composite variables. Otherwise the difference between the x- and y-series will occur in the objective function of certain composite variables more than once. This will lead to bias; the absolute difference between the x- and y-series will be relatively small for those variables, which play a role in the determination of the composite variables. A good example is the *total consumption*. This variable plays a role in determining the *domestic sales*, the *total sales*, *gross production*, *gross* and *net value added* (both *market prices* and *factor costs*), and the (*net*) *operating surplus*. Then, a difference between the original x-value and the generated y-value of total consumption will have some influence ten times. In a similar way, a difference between the x- and y-value of *depreciation* will only have some influence on itself and the *net operating surplus*. Therefore, the influence on the objective function of a difference concerning *total consumption* will be five times as large as a difference concerning *depreciation*. This is why the first part of the objective function will be neglected in the case of composite variables.

Similarly, as results for one size class are dependent on the two others (because the sum of the three size classes should equal 100%), one size class should be neglected in the first part of the objective function as well.

4 Some characteristics of the dataset

This chapter briefly describes the dataset in terms of variables, sectors of industry, and enterprise size classes.

Variables

The variables in the system can be divided into three groups: the exploitation account (28 variables, of which 13 elementary variables), employment (6 variables, of which 4 elementary variables) and investments (1 variable, which is an elementary one). The connections between individual variables are the definitions. For example, *turnover* equals the sum of *total sales* and the *purchase value of merchandise*. In turn, *total sales* are determined by the sum of the sales of four distinguished categories. Regarding these variables, only the sales per category and the *purchase value of merchandise* are elementary (their values being determined in the optimization problem), *total sales* and *turnover* are composite variables. The development of composite variables is taken into account when the objective function is determined (via the second term), but their values are calculated by using their definitions. As an extra constraint it is imposed that *gross value added* (*market prices*, the sum of the *net operating surplus*, *wages*, *employer's social contributions*, *depreciation* and the *net indirect taxes*) may not be negative. This is not a logical constraint, but it is implausible that at the sector/size class level negative gross value added at market prices may be negative.

Changes in inventories, the *net interest payments* and the *net special income and charges* are ignored during calculations. These variables are net variables (for instance, net interest payments is the balance between interest payments and interest received), which are relatively irrelevant with regard to the totals. Moreover, by their nature these variables have an irregular development.

Table 1 Variables

Variable	*Code*	*Calculation*	*Comment*
Exploitation account			
Turnover	omz	afz + iht	
Purchase value of merchandise	iht		Non-negative by definition
Exports (of goods and services)	btx		Non-negative by definition
Total consumption (by households and government)	cgt		Non-negative by definition
Investment sales	ibb		Non-negative by definition
Total intermediate sales	got		Non-negative by definition
Domestic sales	baf	cgt + ibb + got	
Total sales	afz	baf + btx	
Stock changes	nxx		Exogenous
Gross production	bpr	afz + nxx	
Use of commodities	vgh		Non-negative by definition
Use of energy	vet		Non-negative by definition
Use other categories	vov		Non-negative by definition
Total use of intermediate goods and services	vbr	vgh + vet + vov	
Gross value added (market prices)	ybm	bpr − vbr	Non-negative by assumption
Total indirect taxes	ibx		Non-negative by definition
Total subsidies	sux		Non-negative by definition
Total net indirect taxes	ixn	ibx − sux	
Gross value added (factor costs)	ybf	ybm - ixn	
Depreciation	dbx		Non-negative by definition
Net value added (factor costs)	ynf	ybf − dbx	
Wages	lbb		Non-negative by definition
Employer's social contributions	soc		Non-negative by definition
Total labor cost	lbt	lbb + soc	
Net operating surplus	neo	ynf − lbt	
Net interest payments	ren		Exogenous
Net special income and charges	blb		Exogenous
Profit before taxes	win	neo − ren − blb	
Employment			
Labor volume employees	alx		
Labor volume self-employed	azx		
Total labor volume	alz	alx + azx	
Number of employees	wlx		
Number of self-employed	wzx		
Total number of workers	wlz	wlx + wzx	
Investments			
Investments by destination	jbb		

Source: EIM.

Sectors of industry

The twenty sectors of industry distinguished are listed in Table 2.

Table 2 Sectors of industry

Sector	*NACE Rev. 1.1*
Agriculture, hunting and forestry; fishing	A and B
Mining and quarrying	C
Manufacture of food products, beverages and tobacco	DA
Manufacture of basic metals and fabricated metal products	DJ
Manufacture of chemicals, chemical products and man-made fibres; manufacture of rubber and plastic products	DG, DH
Manufacture n.e.c.	DN
Electricity, gas and water supply	E
Construction	F
Sale, maintenance and repair of motor vehicles and motorcycles	50
Wholesale trade	51
Retail trade	52
Hotels and restaurants	H
Transport	60 -63
Communication	64
Financial intermediation	J
Real estate activities	70
Computer and related activities; research and development	71 -73
Other business activities	74
Health and social work	N
Public administration and defence; compulsory social security	L, M

Source: EIM.

Enterprise size classes

The enterprise size classes are listed in Table 3. Note that the traditional Dutch classification of enterprise sizes is used; this classification is only to a limited extent consistent with the EU SME-definition.

Table 3 Enterprise size classes

Size classes	*Code*	*Comment*
Small enterprises	kb	0 up to and including 9 occupied persons
Medium-sized enterprises	mb	10 up to and including 99 occupied persons
Large enterprises	gb	100 or more occupied persons
Total	ab	kb + mb + gb

Source: EIM.

5 Illustration

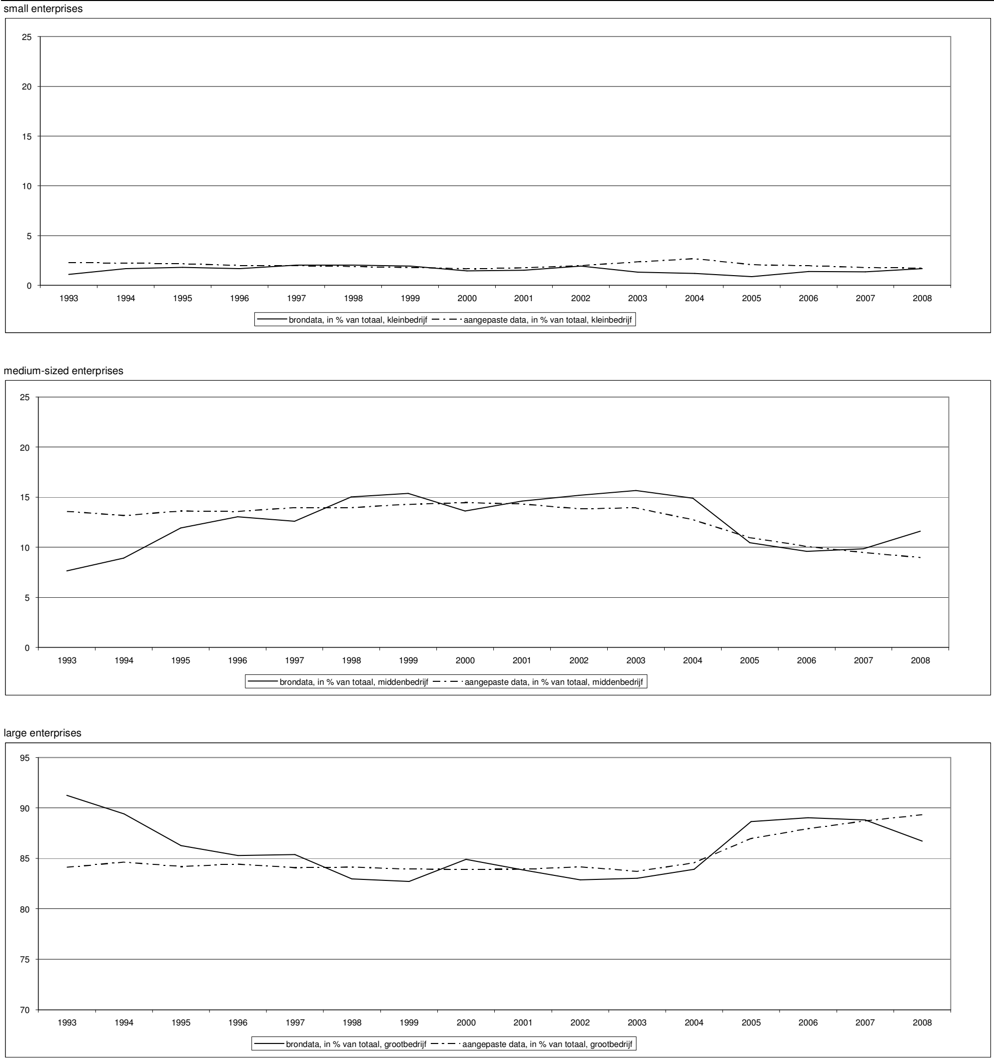
In this chapter the methodology developed in chapter 3 will be applied to two sectors of industry, the chemical industry and the retail trade sector. These sectors are chosen because of their large difference in scale size (the chemical industry is large-scaled, retail trade is small-scaled) and structure (capital-intensive versus labor-intensive; oriented towards exports and sales of intermediate goods and services versus oriented towards sales of consumption goods).

5.1 Application to the chemical industry

Regarding turnover, most remarkable are the rather constant shares of medium-sized and large enterprises in the early years: whereas the source data present a growing share of mediums-sized enterprises at the expense of large enterprises during 1993 -2000, the adjusted series show rather stable market shares (Figure 1). The reason is the rather volatile pattern of value added, for which the source data show a strong increase for large enterprises after 2000 (Figure 4). Smoothing this requires the value added share of large enterprises to fall not too fast before 2000, and because turnover and value added are kinked though definitions, this implies a rather stable development of large enterprises' share in turnover as well. Results for employment (Figure 3) and investment (Figure 4) show that irregular developments a nicely smoothed out, leaving trends unchanged.

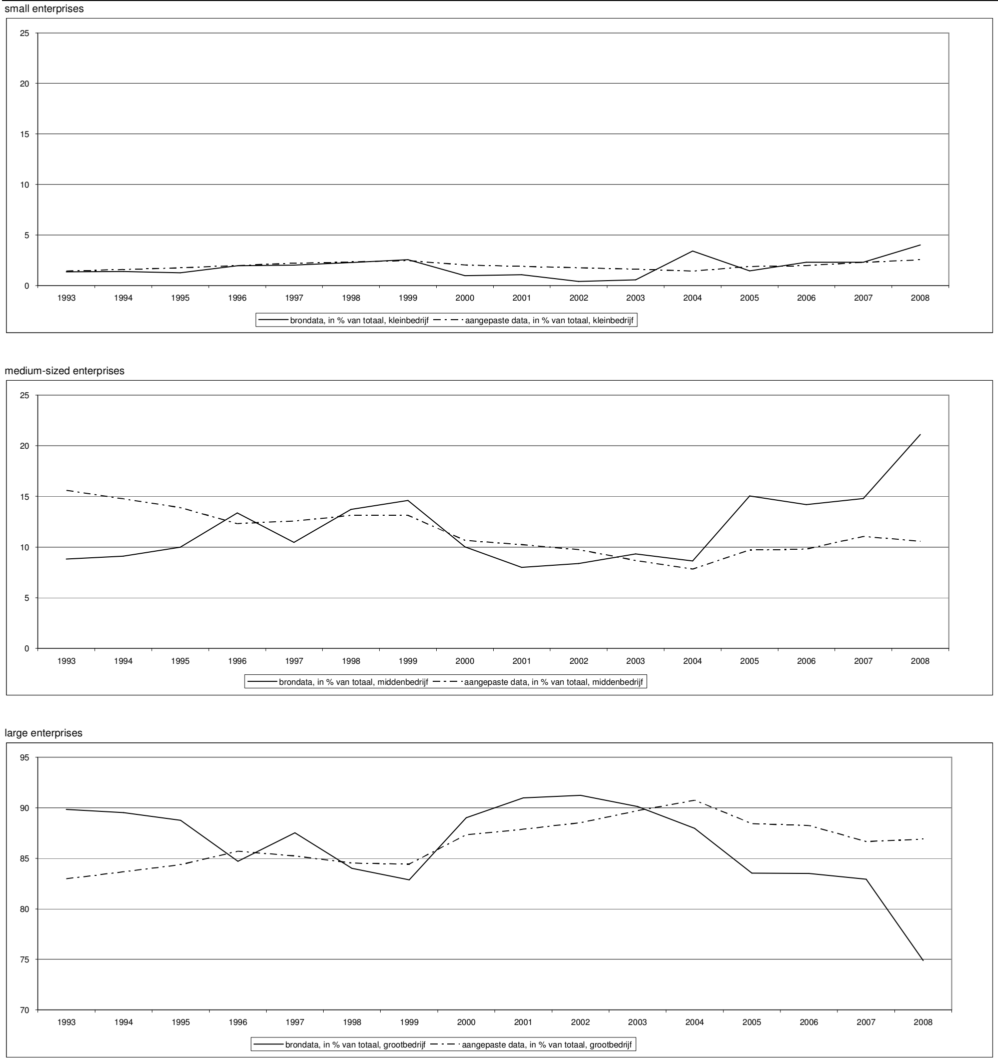
From discussions with industry experts it has been concluded that the adjusted series do better describe structural development patterns than the source data.

Figure 1 size classes' shares in turnover, chemical industry, source data and adjusted series



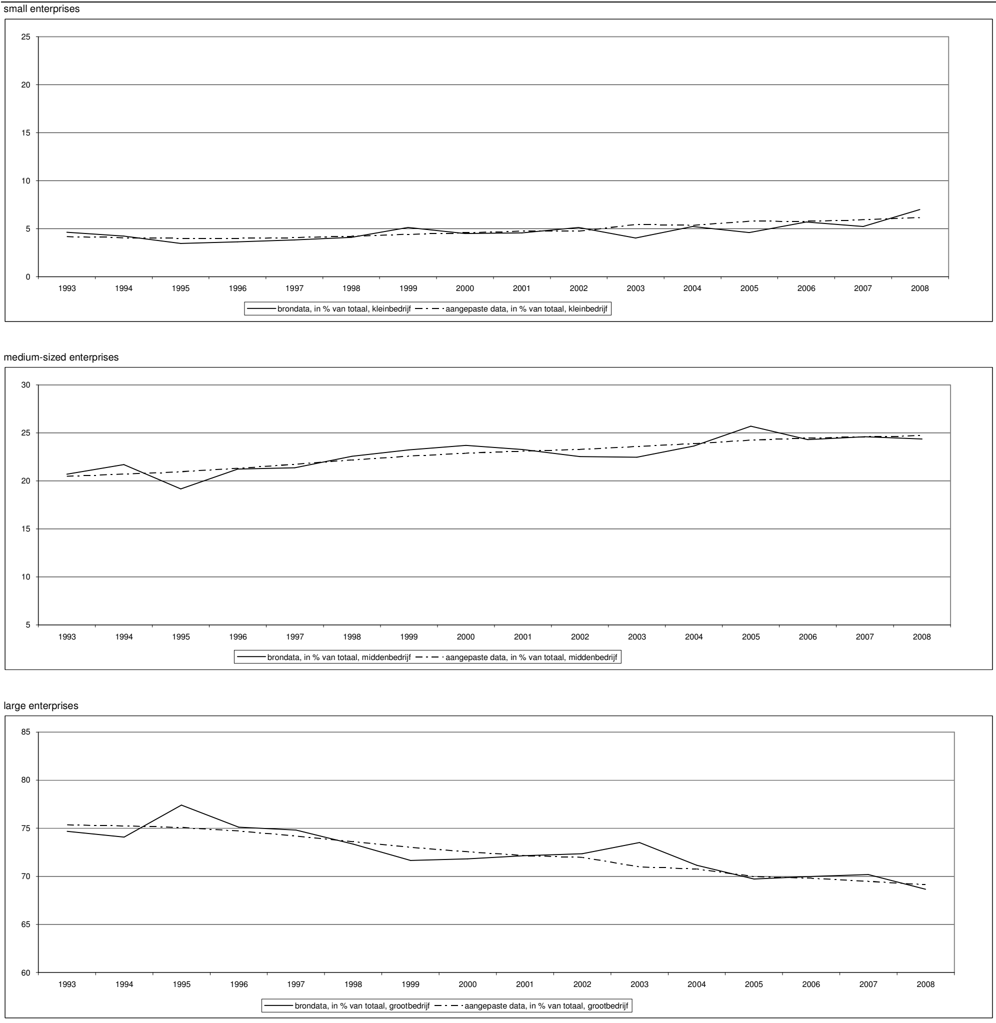
Source: EIM.

Figure 2 size classes' shares in gross value added (factor costs), chemical industry, source data and adjusted series



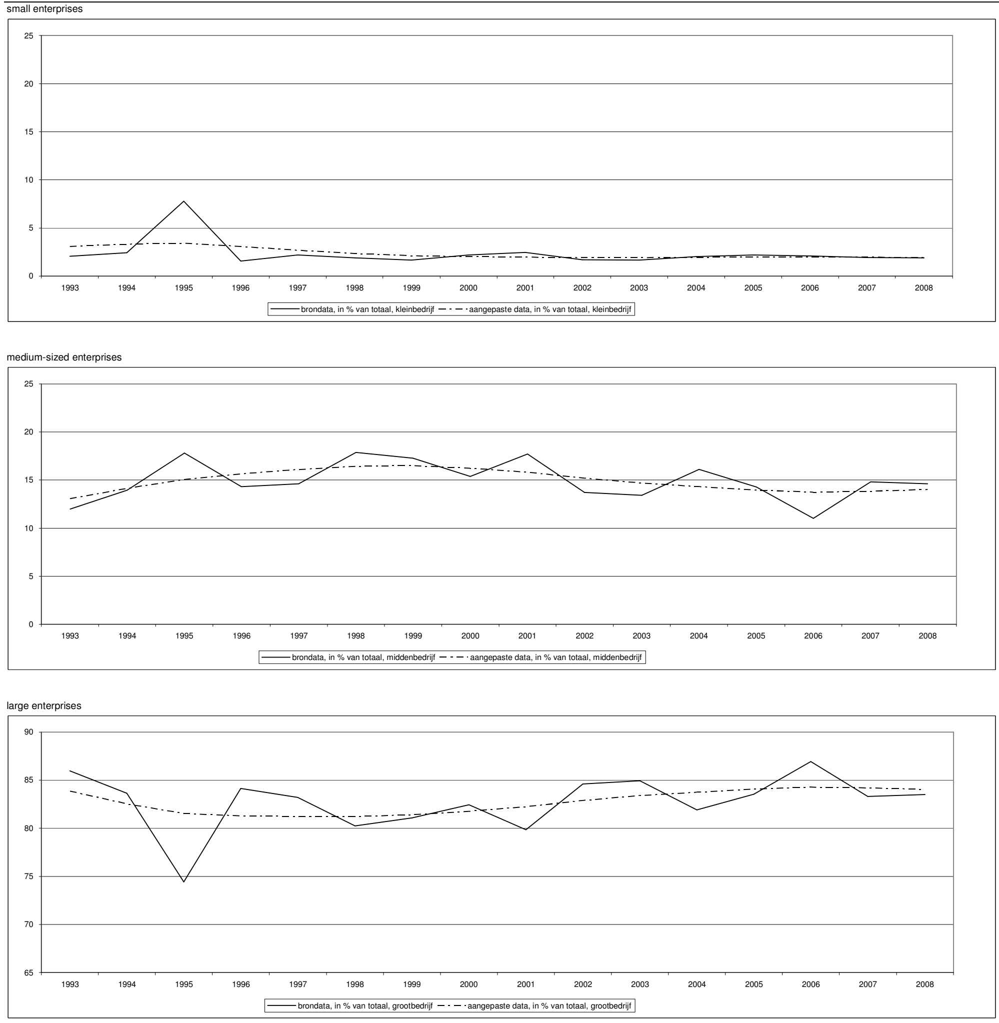
Source: EIM.

Figure 3 size classes' shares in total labour volume, chemical industry, source data and adjusted series



Source: EIM.

Figure 4 size classes' shares in gross investments, chemical industry, source data and adjusted series



Source: EIM.

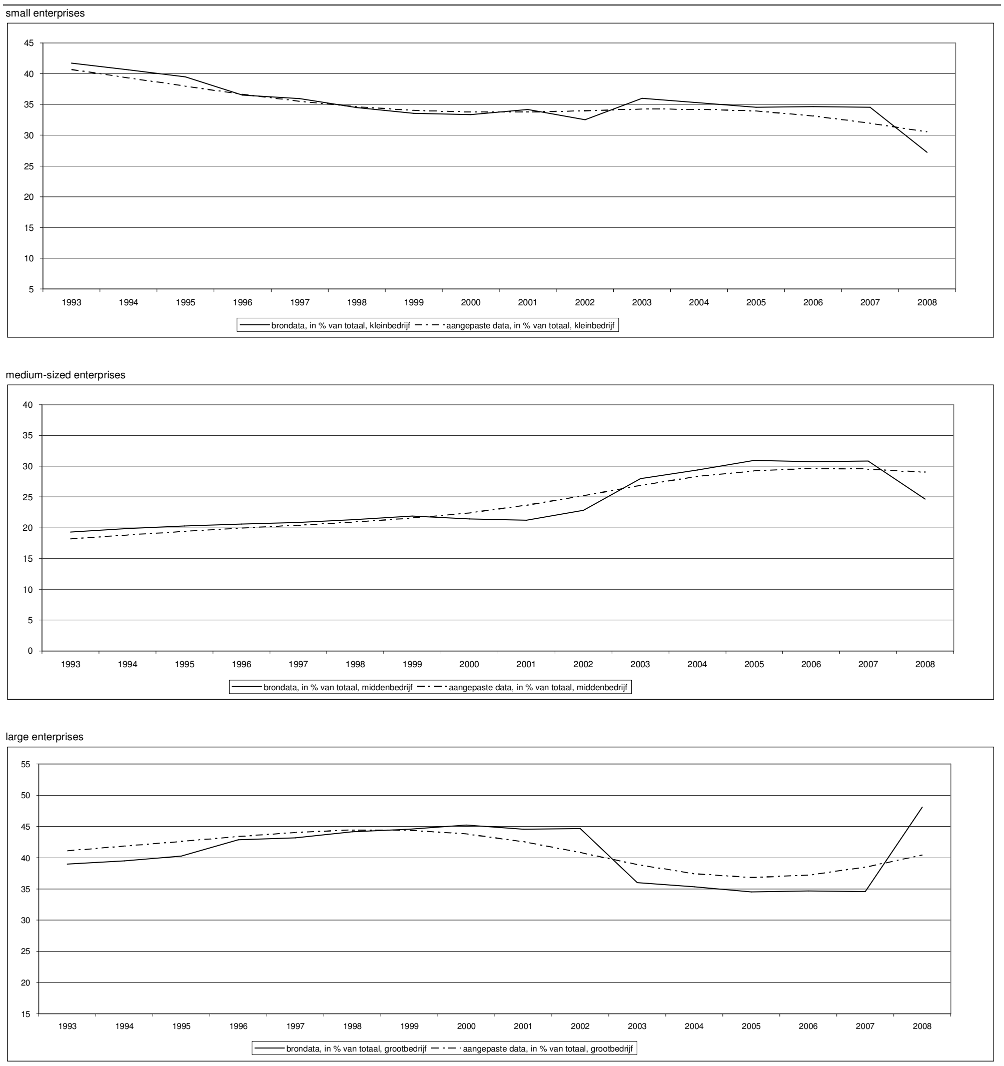
5.2 Application to retail trade

Source data on turnover show an abrupt dip in the share of large enterprises between 2003 and 2008, at the detriment of both small and medium-sized enterprises. The adjusted series show a far more smooth development (Figure 5). Conversely, source data show a large spike in the share of small enterprises in value added in 2000 and 2001, which gradually comes back to a 'normal' level between 2002 and 2005. This is mirrored in the shares of medium-sized and large enterprises. The adjusted series show a more gradual development (Figure 6). The value added/turnover ratio shows a much more realistic development in the adjusted as compared with the source data.

Regarding employment the adjusted series clearly reflect trends already observed in the source data (Figure 7).

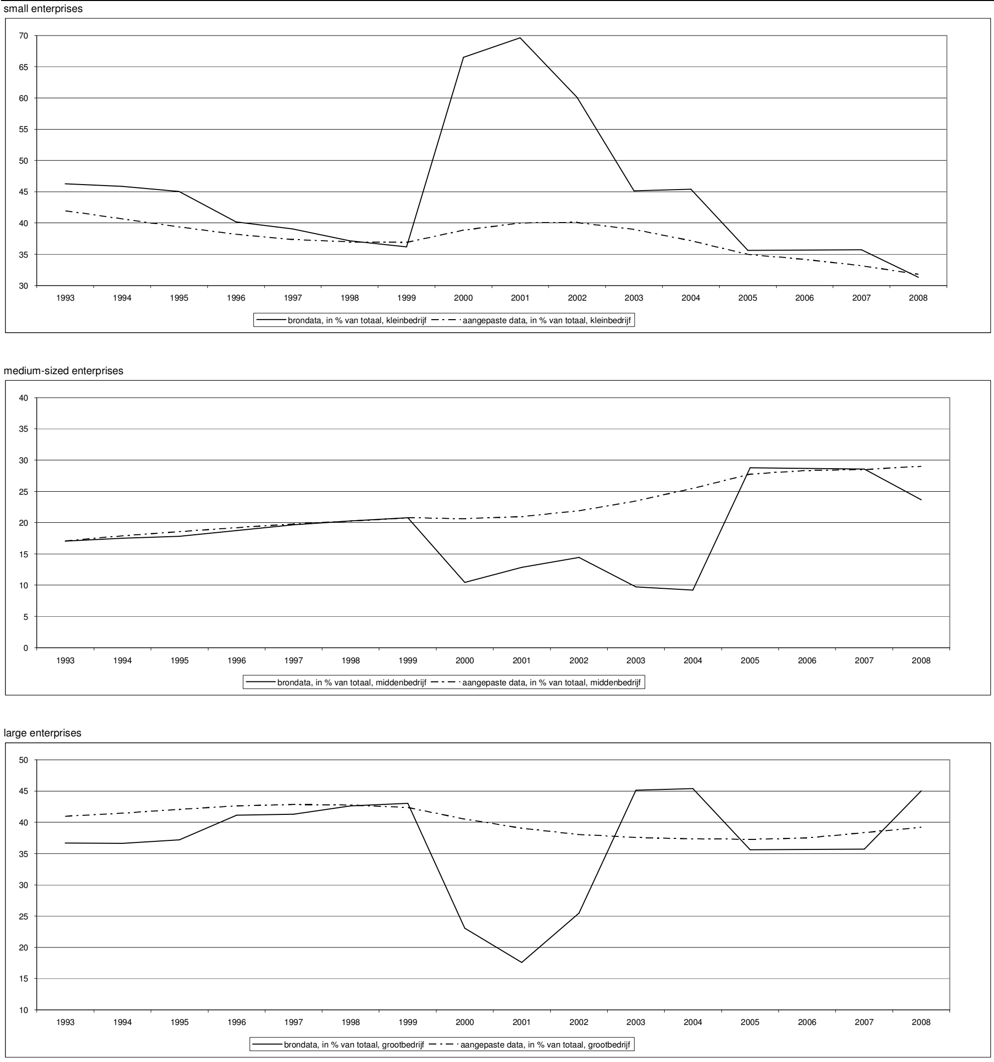
The source data about investments show strong fluctuations in the shares of SMEs and large enterprises. In particular, the fluctuations for small enterprises are not realistic. The adjusted series show a more stable development, with a gradual increase in the share of SMEs until approximately 2003, and a slight reversal of this trend afterwards.

Figure 5 size classes' shares in turnover, retail trade sector, source data and adjusted series



Source: EIM.

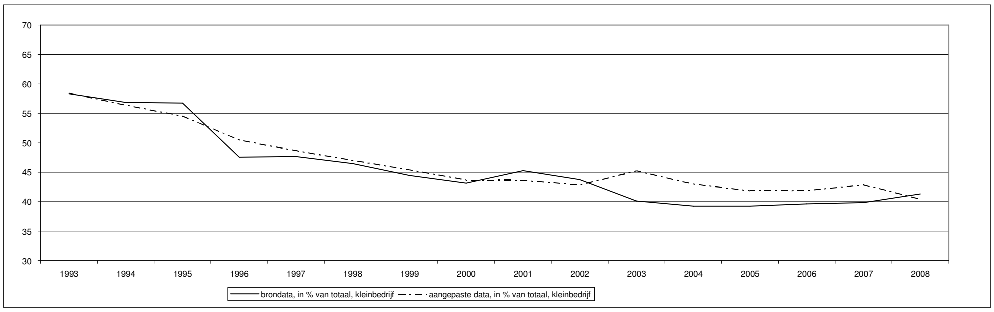
Figure 6 size classes' shares in gross value added (factor costs), retail trade sector, source data and adjusted series



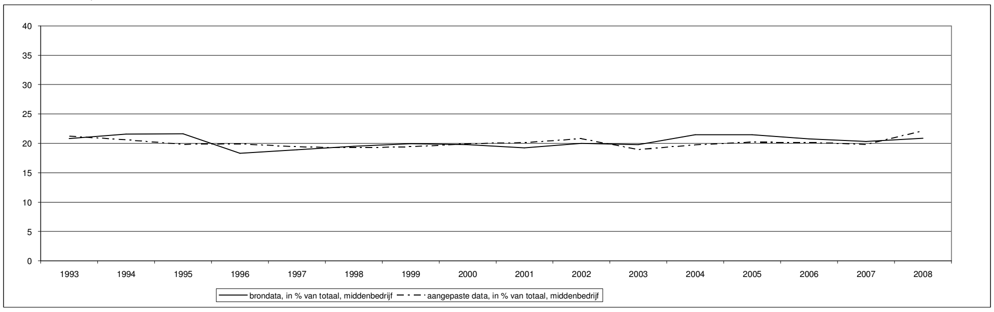
Source: EIM.

Figure 7 size classes' shares in total labour volume, tail trade sector, source data and adjusted series

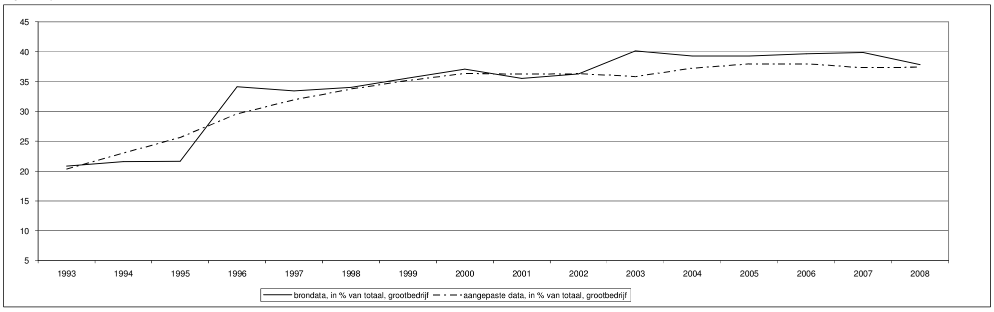
small enterprises



medium-sized enterprises

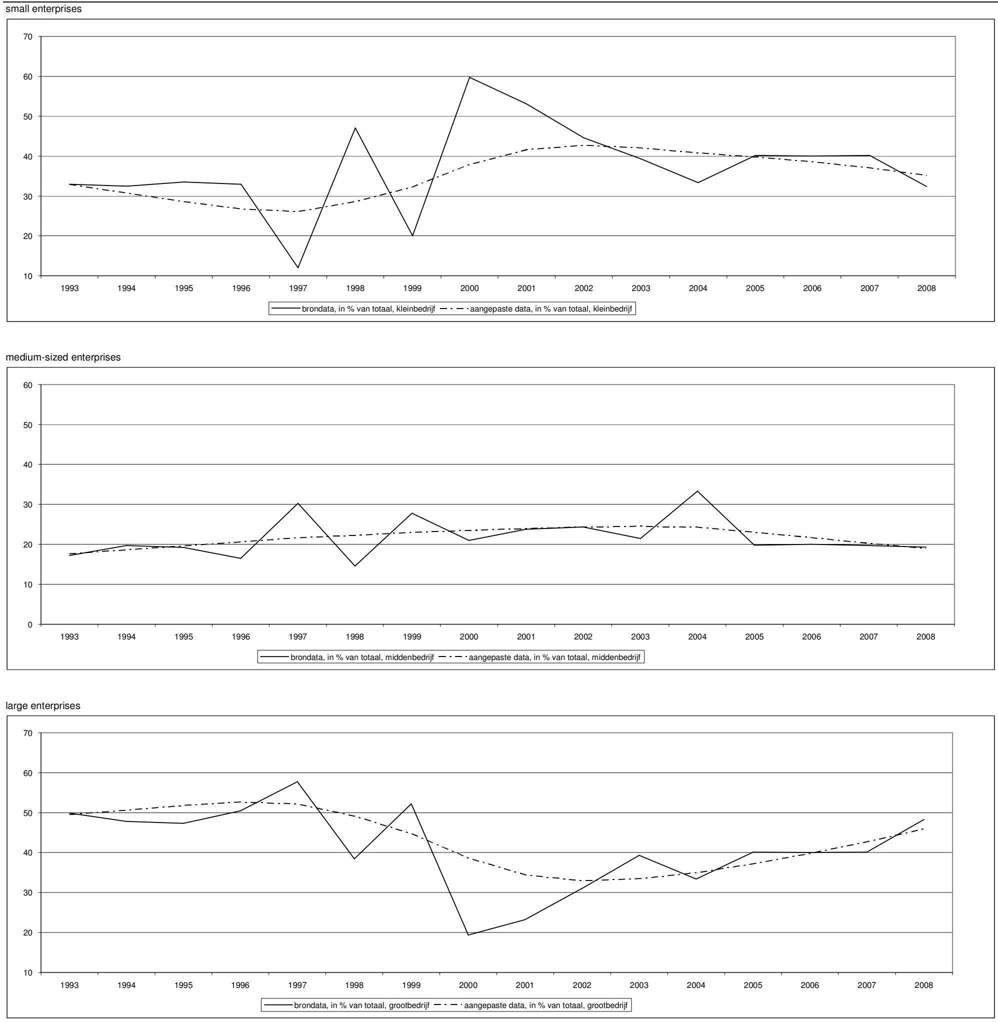


large enterprises



Source: EIM.

Figure 8 size classes' shares in gross investments, retail trade sector, source data and adjusted series



Source: EIM.

Appendix I

Notation:

- $x_{g,i,t}$ is the share of size class g in the sector total for variable i in year t according to the source data.
- $y_{g,i,t}$ is the share of size class g in the sector total for variable i in year t after adjustment.
- Variable set I consists of two subsets: the elementary variables E and the composite variables A. Thus, I= {E, A}, and $x_{g,i\in A,t}$ and $y_{g,i\in A,t}$ are a (linear) combination of the variables $x_{g,i\in E,t}$ and $y_{g,i\in E,t}$ respectively.

The objective function is as follows (the index g is omitted for convenience):

$$\text{(I.1)} \quad D_i = \Delta_i \cdot \Sigma_T (x_{i,t} - y_{i,t})^2 \qquad \text{goodness of fit}$$
$$+ \ \lambda \cdot \Sigma_T \{(y_{i,t} - y_{i,t-1}) - (y_{i,t} - y_{i,t-1})\}^2 \qquad \text{smoothness}$$

If the HP filter was only applied on elementary variables, then the optimization problem would be as follows:

$$\underset{y_{i\in E,t}}{\text{min!}} \quad \Sigma_{I\in E}\, D_i$$

subject to

(1) $y_{g,i,t} \geq 0 \ \forall \ g, i, t$

(2) $\Sigma_G \ y_{g,i,t} = 100 \ \forall \ i, t$

Then, the following applies: Δ_i= 1 en λ= 7.

In the used model, in which all variables are considered interdependently, the optimization problem is as follows:

$$\underset{y_{i\in E,t}}{\text{min!}} \quad \Sigma_{I\in E,A}\, D_i$$

Subject to

(1) $y_{g,i,t} \geq 0 \ \forall \ g, i, t$

(2) $\Sigma_G \ y_{g,i,t} = 100 \ \forall \ i, t$

Δ_i= 1 for elementary variables ($i \in E$) and Δ_i= 0 for composite variables ($i \in A$). For all variables λ= 7 has been assumed[9].

Solutions are iteratively generated using a numerical algorithm. As this is a quadratic optimization problem, there is a unique minimum of the goal function (if it exists), even though there is no guarantee that this is achieved with a unique set of the y_i's. However, it is unlikely that alternative solutions to the optimization problem will lie far apart, as they have to fulfill the condition of being 'close' to the source data x_i.

[9] The same smoothing parameter value applies to all variables to avoid that the smoothness for elementary variables gets a smaller or greater weight than the smoothness of composite variables. By choosing λ= 7 for both elementary and composite variables it is obtained that the smoothness of each variable is equally weighted.

References

Ahumada, H. and Garegnani, M.L. (1999). Hodrick-Prescott Filter in Practice. La Plata: Universidad Nacional de La Plata.

Ahumada, H. and Garegnani, M.L. (2000). Assesing HP Filter Performance for Argentina and US Macro Aggregates. Journal of Applied Economics, 3 (2), 257–284.

Baxter, M. and King, R.G. (1999). Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series. The Review of Economics and Statistics, 81 (4), 575–593.

Beveridge, S. and Nelson, C.R. (1981). A New Approach to the Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'. Journal of Monetary Economics, 7, 151–174.

Burns, A.F. and Mitchell, W.C. (1946). Measuring Business Cycles. New York: National Bureau of Economic Research.

Chamberlin, G. (2006). Fitting trends to time series data. Newport: Office for National Statistics.

Chuan, B.C. (2006). Univariate Detrending Using SAS: Some Examples and Applications (University Information Technology Services, Center for Statistical and Mathematical Computing, Indiana University)

Guay, A. and St-Amant, P. (1997). Do the Hodrick-Prescott and Baxter-King Filters Provide a Good Approximation of Business Cycles?. Montréal: Université du Québec à Montréal.

Harvey, A. and Trimbur, T. (2008). Trend estimation and the Hodrick-Prescott Filter, Tokyo: Journal of the Japan Statistical Society, 38 (1), 41–49.

Hodrick, R.J. and Prescott, E.C. (1997). Postwar U.S. Business Cycles: An Empirical Investigation. Journal of Money, Credit and Banking, 29 (1), 1–16.

Kalekar, P.S. (2004). Time series Forecasting using Holt-Winters Exponential Smoothing. Bombay: Kanwal Rekhi School of Information Technology.

Lasdon, L.S., Waren, A.D., Jaim, A. and Ratner, M. (1978). Design and Testing of a Generalized Reduced Gradient Code for Nonlinear Programming. Journal ACM Transactions on Mathematical Software, 4 (1), 34–50.

LaViola Jr., J.J. (2003). Double Exponential Smoothing: An Alternative to Kalman Filter-Based Predictive Tracking. Providence: Brown University Technology Center for Advanced Scientific Computing and Visualization.

Ley, E. (2006). Knowledge Brief for Bank Staff: The Hodrick-Prescott Filter. Washington DC: The World Bank.

Pommée, M. and Van Dalen, J. (1997). Een input-outputtabel naar grootteklasse van de Nederlandse economie voor 1992: kleine, middelgrote en grote bedrijven. Publicatie M&O.009, Centraal Bureau voor de Statistiek, Voorburg/Heerlen.

Ravn, M.O. and H. Uhlig (2002). On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations. The Review of Economics and Statistics, 84 (2), 371–380.

Trewin, D. (2003). A Guide to Interpreting Time Series – Monitoring Trends. Information Paper No. 1349.0, Australian Bureau of Statistics, Canberra.

Woitek, U. (1998). A Note on the Baxter-King Filter. Glasgow: University of Glasgow.

The results of EIM's Research Programme on SMEs and Entrepreneurship are published in the following series: Research Reports and Publieksrapportages. The most recent publications of both series may be downloaded at: www.entrepreneurship-sme.eu.

Recent Research Reports and Scales Papers

H201203	1-2-2012	Do Small Businesses Create More Jobs? New Evidence for Europe
H201202	19-1-2012	Trends in entrepreneurial Activity in Central and East European Transition Economies
H201201	9-1-2012	Globalization, entrepreneurship and the region
H201119	2-1-2012	The risk of growing fast
H201118	22-12-2011	Beyond Size: Predicting engagement in environmental management practices of Dutch SMEs
H201117	22-12-2011	A Policy Theory Evaluation of the Dutch SME and Entrepreneurship Policy Program between 1982 and 2003
H201116	20-12-2011	Entrepreneurial exits, ability and engagement across countries in different stages of development
H201115	20-12-2011	Innovation barriers for small biotech, ICT and clean tech firms: Coping with knowledge leakage and legitimacy deficits
H201114	20-12-2011	A conceptual overview of what we know about social entrepreneurship
H201113	20-12-2011	Unraveling the Shift to the Entrepreneurial Economy
H201112	24-11-2011	Bedrijfscriminaliteit
H201111	25-8-2011	The networks of the solo self-employed and their success
H201110	23-6-2011	Social and commercial entrepreneurship: Exploring individual and organizational characteristics
H201109	9-5-2011	The relationship between firm size and economic development: The Lucas hypothesis revisited
H201108	22-3-2011	Corporate Entrepreneurship at the Individual Level: Measurement and Determinants
H201107	30-01-2011	Determinants of high-growth firms
H201106	13-1-2011	Determinants of job satisfaction across the EU-15: A comparison of self-employed and paid employees
H201105	13-1-2011	Gender, risk aversion and remuneration policies of entrepreneurs
H201104	11-1-2011	The relationship between start-ups, market mobility and employment growth: An empirical analysis for Dutch regions
H201103	6-1-2011	The value of an educated population for an individual's entrepreneurship success
H201102	6-1-2011	Understanding the Drivers of an 'Entrepreneurial' Economy: Lessons from Japan and the Netherlands

H201101	4-1-2011	Environmental sustainability and financial performance of SMEs
H201022	16-11-2010	Prevalence and Determinants of Social Entrepreneurship at the Macro-level
H201021	20-10-2010	What determines the volume of informal venture finance investment and does it vary by gender?
H201019	9-8-2010	Measuring Business Ownership Across Countries and Over Time: Extending the COMPENDIA Data Base
H201018	6-7-2010	Modelling the Determinants of Job Creation: Microeconometric Models Accounting for Latent Entrepreneurial Ability
H201016	29-4-2010	The Nature and Prevalence of Inter-Organizational Project Ventures: Evidence from a large scale Field Study in the Netherlands 2006-2009
H201015	27-4-2010	New Firm Performance: Does the Age of Founders Affect Employment Creation?
H201014	1-4-2010	Van defensief MKB-beleid naar offensief ondernemerschapsbeleid
H201013	16-3-2010	Human capital and start-up succes of nascent entrepreneurs
H201012	16-3-2010	New business creation in the Netherlands
H201011	16-3-2010	Factors influencing the entrepreneurial engagement of opportunity and necessity entrepreneurs
H201010	16-3-2010	The more business owners the merrier? The role of tertiary education
H201009	2-3-2010	Open, distributed and user-centered: Towards a paradigm shift in innovation policy
H201008	1-2-2010	Geographical distance of innovation collaborations
H201006	21-1-2010	Family ownership, innovation and other context variables as determinants of sustainable entrepreneurship in SMEs: An empirical research study
H201005	12-1-2010	Intrapreneurship – An international study
H201004	12-1-2010	Schumpeter versus Kirzner: An empirical investigation of opportunity types
H201003	7-1-2010	Institutions and entrepreneurship: The role of the rule of law
H201002	5-1-2010	Patterns of innovation networking in Dutch small firms
H201001	4-1-2010	Administratieve lasten en ondernemerschap
H200911	29-1-2010	Ambitious entrepreneurship, high-growth firms and macroeconomic growth
H200910	4-2-2010	Entrepreneurial exit and entrepreneurial engagement
H200909	10-3-2009	Entrepreneurship Education Monitor (EEM)
H200908	3-3-2009	Internationale samenwerking door het Nederlandse MKB
H200907	2-3-2009	The Dynamics of Entry and Exit