Discussion paper

From quarterly to monthly turnover figures using nowcasting

Daan Zult, Sabine Krieg, Bernd Schouten, Pim Ouwehand and Jan van den Brakel

March 2020

Content

Summary
Short term business statistics for the industries Services and Transportation in the Netherlands are compiled from Value Added Tax (VAT) administrations. In the past business where required to declare VAT at a monthly frequency, but since 2009 the legal regulation in the Netherlands allows companies to decide for themselves to file a tax return on a monthly, quarterly or even at an annual basis. As a result, a substantial amount of businesses declare VAT at a quarterly frequency and therefore the VAT-based short term statistics in the Netherlands are currently published on a quarterly frequency. According to a Eurostat regulation, short term business statistics have to be published on a monthly frequency from 2021 on. In this paper several estimation procedures are compared to compile timely monthly short term business statistics from VAT declarations from companies reporting partly on a monthly basis and partly on a quarterly basis. To produce monthly figures for periods after obtaining all available monthly and quarterly information a temporal disaggregation method proposed by Chow Lin is proposed. For timely estimates of the months during the last quarter, the temporal disaggregation method of Chow Lin is extended with a nowcast method based on a bridge model.

Keywords
Short-term business statistics, now casting, temporal disaggregation

1. Introduction

Short term business statistics are important indicators about the short term development of the economy. Their aim is to describe the development of turnover in economic sectors on different levels of aggregation with respect to both activity and company size.

In the Netherlands, short term business statistics are mainly derived from value added tax (VAT) administrations for the industries concerning Services and Transportation. Enterprises declare VAT on different frequencies. A small number of businesses declare turnover yearly, but the major part declares turnover on quarterly or monthly basis. The current publication frequency of short term statistics is quarterly. However, a new European regulation, called the Framework Regulation Integrating Business Statistics (FRIBS), will require Statistics Netherlands to also publish part of the business statistics on a monthly frequency. Statistics on a monthly frequency cannot be compiled directly from the VAT records, because a substantial part of the population declares VAT at a lower frequency. Therefore, the question addressed in this paper is how to derive short term statistics on a monthly frequency using these administrative data where monthly data is only available from a relatively small and possibly selective part of the population.

Monthly figures could be produced by applying primary data collection. Although primary data collection will always be necessary for certain specific groups of enterprises, it is not in line with Statistics Netherlands' policy to collect this information through surveys, since Statistics Netherland also has to reduce the administrative costs and response burden for enterprises. For this reason, Statistics Netherlands has started to develop some methods to produce monthly figures based on the VAT administration. In Daalmans (2015), a Denton approach is developed, and in Van den Brakel and Krieg (2016) a structural time series model is developed. In this paper, both methods are further developed and compared. Furthermore, the adjustment method of Chow and Lin (1971) and a nowcast method based on bridge equations (Baffigi et al., 2004) are investigated.

The first estimate of the monthly index series must be computed 60 days after the reference month according to European regulation. When the month concerns the first or second month of a quarter, no information about the current quarter of the quarterly declarants is yet available, and the estimate is therefore based on current information about the monthly declarants only, combined with information from the past from both monthly and quarterly declarants.

The objective of producing a monthly index series can therefore be interpreted as a combination of a nowcasting problem and a temporal disaggregation problem. Nowcasting is defined as estimating parameters in real time, i.e. making preliminary estimates during the reference period with incomplete data about this period. The difference with regular forecasting is that some information about the period of interest is already available. In our case, we have the monthly VAT-series up to the month we want to nowcast and the quarterly index up to the last quarter before that month. Specific methods exist to utilize this data structure. The method of Daalmans (2015) takes these specific aspects into account, where an ARIMA model is applied to nowcast the ratio of the auxiliary monthly series (based on the monthly declarants) and the total series (based on monthly and quarterly declarants). Structural time series models as developed in Van den Brakel and Krieg (2016) can deal with the problem of nowcasting by considering the information of the quarterly declarants as missing.

Temporal disaggregation means that the series of the monthly declarants is used as auxiliary information to estimate the monthly seasonal pattern of the quarterly declarants.

In Section 2 we take a closer look at the data we use as input. We discuss various properties of the data sources and the way the resulting time series are constructed. In the next section we discuss the various estimation methods considered in this research. Both ARIMA-based models and structural time series models will be considered. In Section 4 we present our results for the various methods. Finally, Section 5 gives our conclusion and some suggestions for further research.

2. Data for Short Term Statistics

Statistics Netherlands publishes short term indicators on different levels of detail. The most detailed level are primary publication cells (PPC's), but there are also publications on higher levels. The PPC's are based on the NACE-code ("Nomenclature statistique des activités économiques dans la Communauté européenne", which is a classification based on the economic activity of the enterprises) and almost correspond one to one with the classes of the NACE-code at the four digit level. Other important publication domains are based on the size class of the enterprise, which is based on the number of working persons. In this paper, we focus mostly on publications on the level of PPC's.

In order to produce the statistics, two data sources are distinguished. For the majority of the enterprises VAT is used, as already discussed in the introduction. For a small group containing the largest and most complex enterprises (called Top-W), Statistics Netherlands maintains primary data collection on a quarterly frequency. These enterprises are very important for the estimation of the official economic quarterly index series. The structure of these enterprises is complex, and their NACE-code and size class can change quite often, for example due to mergers. This makes this subpopulation very dynamic.

There are several problems when using the tax register in making statistics. First, the fiscal unit for which the turnover is known, is not always exactly the statistical unit of interest. Second, not for all fiscal units the turnover is known, for example because the enterprise has a tax exemption, or because the enterprise is late in the tax return. These two problems are relatively small and acceptable solutions are available for the computation of quarterly figures. In this paper we assume that these solutions are also acceptable when monthly figures are computed. A third problem concerns the period in which enterprises fill in their tax return. Since 2009, they are generally free to choose how frequent they fill out their tax return, i.e. annually, quarterly or monthly. In practice, the majority chooses a quarterly tax return. Therefore, quarterly index series can be constructed easily from the VAT data, but because the share of monthly declarants is much smaller, the construction of monthly index series is not straightforward.

A first approach to compute a monthly index series is by assuming that the enterprises with monthly tax return are a random sample from the population of all enterprises. Then the monthly index series can be computed by weighting the monthly tax payers to the entire population, using for example size class as auxiliary information. This method is called pseudo design-based estimation in the literature (Baker et al., 2010). It is, however, not exactly known why enterprises choose a specific frequency of tax return, and it is likely that the assumption about the randomness of the sample does not hold. It can therefore be anticipated that simple weighting schemes based on e.g. only size class are not sufficient to correct for this selectivity, since these methods assume that the monthly declarants within each subpopulation are representative for the quarterly declarants in the corresponding subpopulation. As an alternative, selection bias might be reduced with model-based procedures like the methods of Daalmans (2015) and Van den Brakel and Krieg (2016) that attempt to combine the VAT information observed on a monthly frequency with the VAT observed at quarterly frequency.

2.1 Quarterly index series

Statistics Netherlands does not publish the sum of the turnover of the enterprises, but an index, based on this turnover. In this paper we write this quarterly index as a monthly series:

$$y_t^Q = \begin{cases} \text{quarterly index} & \text{for } \frac{t}{3} \in \mathbb{N}, i.e.\ t = 3,6,\dots,T_Q \\ \text{NA} & \text{for } \frac{t}{3} \notin \mathbb{N}, i.e.\ t = 1,2,4,5,7,\dots,T_Q - 1, \end{cases} \quad (1)$$

where $\mathbb{N}$ is the set of natural numbers and T_Q is the length of the series in months until the third month of the last available quarter. Each quarter the quarterly declarants declare their turnover at $t = 3,6,\dots,T_Q$ and the monthly declarants at $t = 1,2,3,\dots,T_M$ where T_M can be either T_Q, $T_Q + 1$ or $T_Q + 2$. In order to obtain the quarterly turnover, for each quarter, the turnover of the monthly declarants are added to the turnovers of the quarterly declarants. In order to obtain an index, a standard calculation would be z_t^Q / z_3^Q. However, unfortunately this does not describe the real development. For instance when a company merges with another company from a different PPC, this leads to jumps in the index that do not reflect real economic developments but are induced by administrative changes. Therefore it is agreed internationally to exclude such enterprises from the index, in the period that an administrative change took place, even though this may often concern enterprises of substantial size. The decision whether or not to exclude an enterprise is made by economic experts at the department of business statistics. Finally, we should note that new and 'dying' enterprises are included in the computation, with a turnover of 0 in one of the periods.

With this administrative reality in mind, each period the declarants are identified that also declared in the previous quarter and the sum of their turnovers is defined as $\vec{z}_{t-3}^{Q}$and $\overleftarrow{z}_t^{Q}$ for the previous and current quarter. This implies that $\vec{z}_t^{Q}$ is the turnover of enterprises in period $\{t-2, t-1, t\}$, who also declared tax in period $\{t+1, t+2, t+3\}$ and $\overleftarrow{z}_t^{Q}$ the turnover of enterprises in period $\{t-2, t-1, t\}$, who also declared tax in period $\{t-5, t-4, t-3\}$. At $t = 3$ the index y_3^Q is set at 100. Then a preliminary quarterly index is computed as:

$$y_t^Q = y_{t-3}^Q \frac{\overleftarrow{z}_t^{Q}}{\vec{z}_{t-3}^{Q}}, \qquad \text{for } t = 3,6,\ldots,T_Q. \tag{2}$$

When the preliminary index series is computed, the complete series is rescaled to make sure that the mean of one specific year (the base year) is equal to 100. At the moment, this is the year 2015 for the officially published series. It is common practice that the base year is shifted every 5 or 10 years and the index series is rescaled. Currently, both the original and seasonally adjusted index series are published at quarterly frequency. This will not change when the monthly index series will be published as well. The seasonal adjustment is outside the scope of this research project. In this paper we will refer to y_t^Qas the published quarterly index series.

2.2 Monthly auxiliary series

The input of the different methods which are applied in this paper are two time series. The first series is the quarterly index series which was discussed in the previous subsection that is available for $t = 3,6,\ldots,T_Q$. The second series is a monthly index series x_t^M that is available for $t = 1,2,\ldots,T_M$, where $T_M \geq T_Q$, because months can be available before the end of the quarter. This monthly index series is based on the turnover information of the enterprises who fill in their tax return on a monthly basis, i.e. $\overleftarrow{z}_t^{M}$ and $\vec{z}_t^{M}$, which are defined similarly as $\overleftarrow{z}_t^{Q}$ and $\vec{z}_t^{Q}$, i.e. $\vec{z}_t^{M}$ is the turnover of enterprises in period t, who also declared tax in period t+1 and $\overleftarrow{z}_t^{M}$ is the turnover of enterprises in period t, who also declared tax in period t-1.
Alternatively one could construct $\overleftarrow{z}_t^{M}$ and $\vec{z}_t^{M}$ a bit differently, by taking background characteristics (for example size class of the enterprises) into account. This is because monthly declarants might have a different index curve from quarterly declarants and background characteristics might be able to correct for this difference. This alternative approach might be part of future research.

In this paper the definition of the monthly auxiliary series is.

$$x_t = x_{t-1} \frac{\overleftarrow{z}_t^M}{\overrightarrow{z}_{t-1}^M}, \text{ for t = 1,2,...,} T_M. \quad (3)$$

where x_t is scaled so that $\frac{(x_1+x_2+x_2)}{3} = y_3^Q$. Unlike $\overleftarrow{z}_t^Q$ and $\overrightarrow{z}_t^Q$, the sums of turnover $\overleftarrow{z}_t^M$ and $\overrightarrow{z}_t^M$ are based on the monthly declarants only, which are not checked manually and adjusted by experts on the monthly level, they are simply based on the raw monthly turnover data. These two factors imply that x_t is unlikely to be a good candidate to describe the monthly development of each PPC.

2.3 Data visualization

In this paper, series of the years 2010 Q1 – 2019 Q2 are considered. We focus on twelve PPC's. The first six PPC's are from the hospitality sector, NACE section I: Hotels and similar accommodation, Holiday and other short-stay accommodation, Restaurants, Mobile food service activities, Event catering and other food service activities, Beverage serving activities. In the continuation of this paper, we will refer to these PPC's as Hotels, Other accommodation, Restaurants, Fastfood, Catering and Pubs. In this paper, we will mainly focus on these six, since they can also be aggregated and have a meaningful interpretation. The other six will be used to check that our findings generalize to other PPC's. They are from sections J, M, N and S: Publishing activities, Legal activities, Accounting, bookkeeping and auditing activities, tax consultancy, Employment activities, Office administrative, office support and other business support activities, Repair of computers and personal and household goods. They will be referred to as Publishers, Legal activities, Accountants, Employment activities, Other Business Support, Repair of household goods.

Figure 2.1a-f show y_t^Q and x_t and the quarterly mean of x_t for the six PPC's in the hospitality sector.

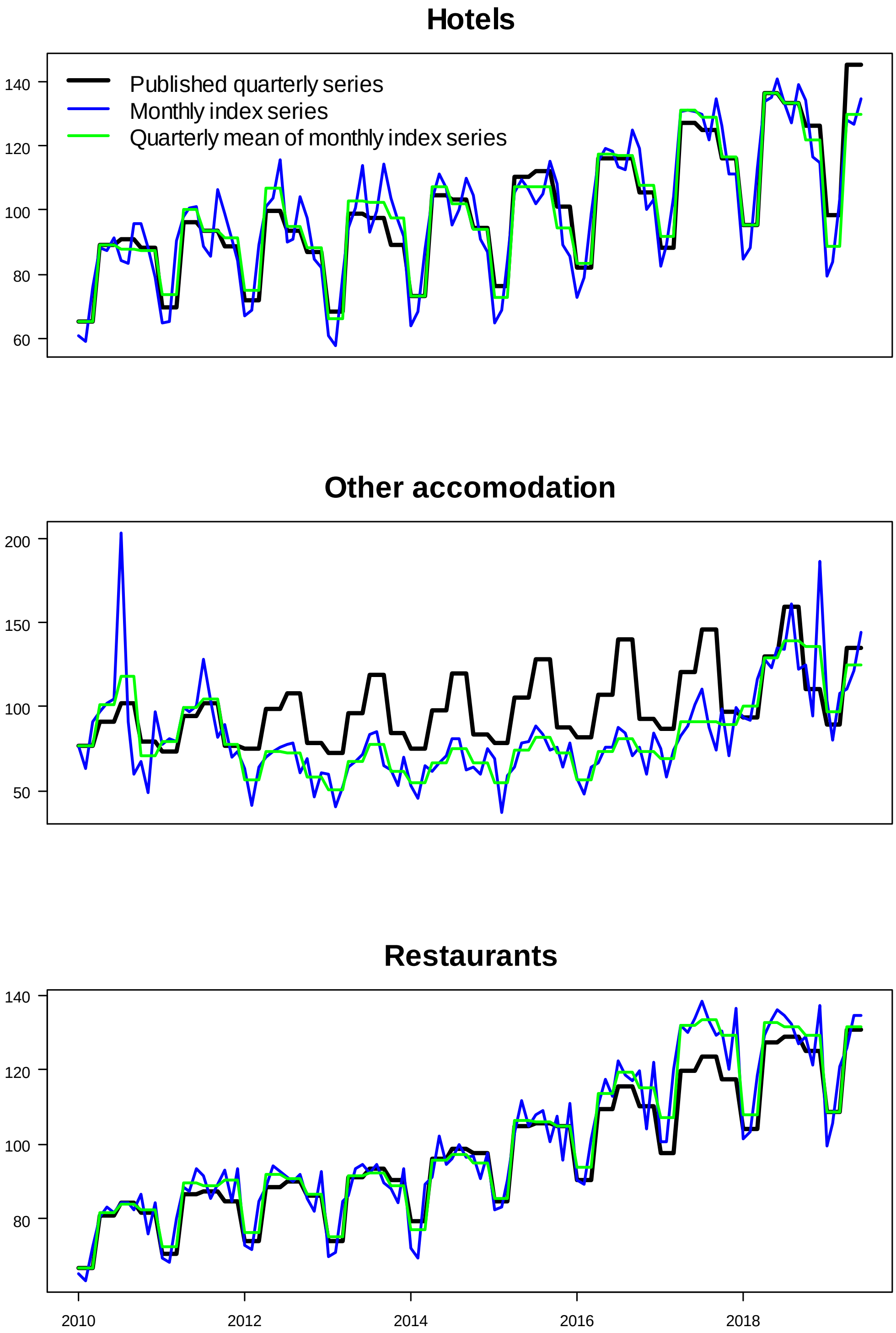


Figure 2.1a-c: monthly and quarterly index series and the quarterly mean of the monthly index series, in the hospitality sector.

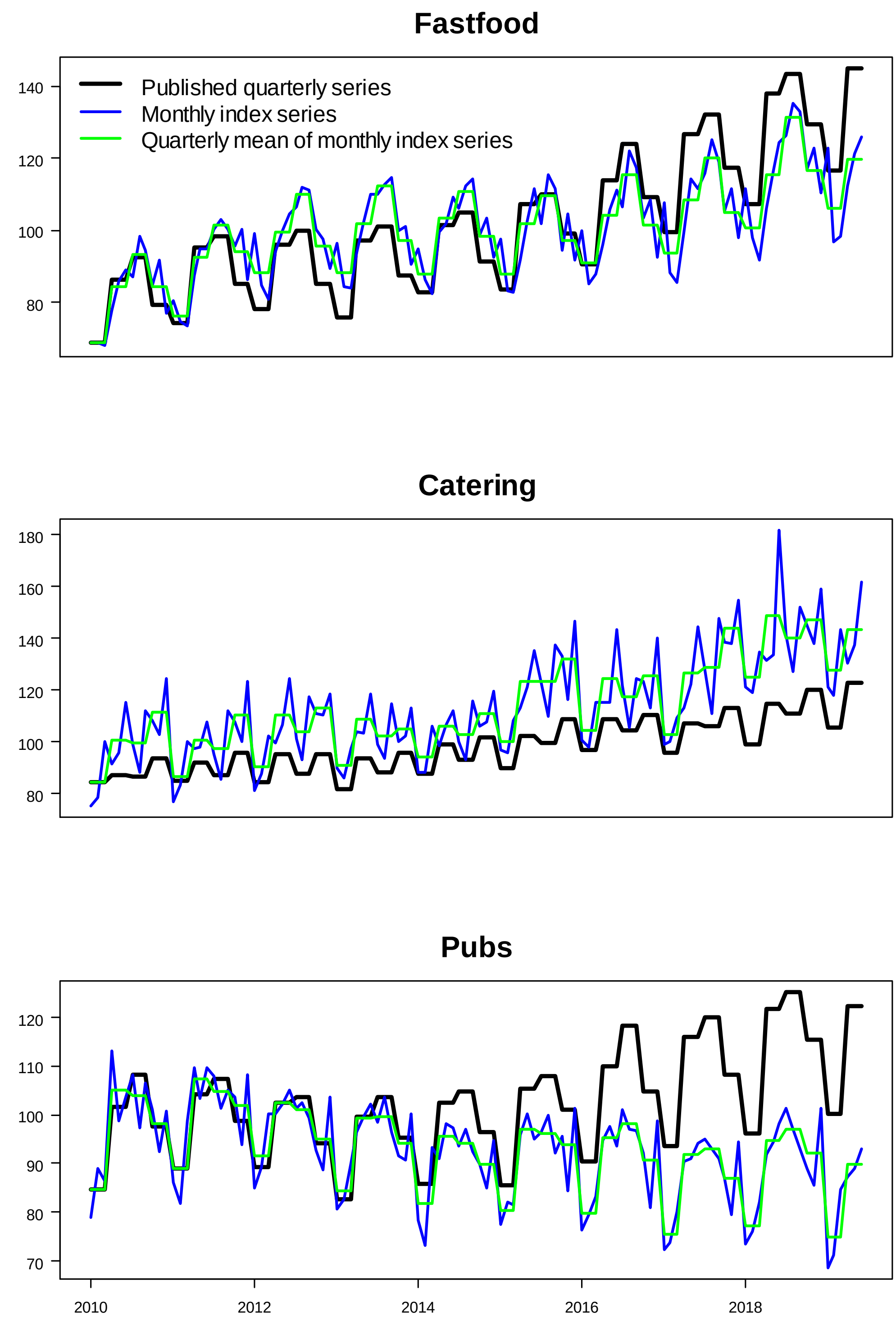


Figure 2.1d-f: monthly and quarterly index series and the quarterly mean of the monthly index series, in the hospitality sector.

When we focus on the graph of Hotels, the trend of both quarterly series (published and quarterly mean of monthly series) is increasing at more or less the same rate over the considered period, with some small differences. For example, the growth of the published series between 2013 and 2014 is larger than the growth of the (quarterly mean of the) monthly series. The seasonal pattern in both quarterly series is quite similar. For Restaurants, the results are similar as for Hotels both for trend and seasonal pattern. For Other accommodation, the (quarterly mean of the) monthly series decreases in the first two years, whereas the quarterly series slightly increases. Furthermore, the seasonal patterns are quite different here. Especially in the first years, the monthly series is quite irregular due to outliers. For Catering, the increase of trend of the published series is much weaker than the one of the quarterly mean of the monthly series. The seasonal pattern are quite similar. Finally, for Pubs, the trends of both series have opposite directions, and the seasonal patterns are again similar. We do not show figures for the six other PPC's here.

For most PPC's, the seasonal pattern of the quarterly index (based on all declarants) is more or less similar to the seasonal pattern of the quarterly mean of the monthly index, based on the monthly declarants (this as also true for the six other PPC's). It is plausible that the monthly seasonal pattern of all declarants (if they all would declare on a monthly basis) would be similar to the pattern of the monthly declarants. It is not possible to check this assumption, but at least, from these figures there is no proof that it is not true. The computation of the monthly index for all declarants is based on this assumption.

The development of each series is determined partly by starting and dying enterprises. When these enterprises are overrepresented in either the monthly or quarterly series, this may lead to substantial differences between both series. In fact, when we look at the data we see that in general most new firms both have higher growth rates and choose to report tax on a quarterly basis, which causes y_t^Q to grow at a faster pace than x_t. However, this phenomenon is not a fixed rule, it could be the other way around for some PPC's and there might be other unobserved differences between the monthly and quarterly declarants, resulting in the differences in the series as shown in Figure 2.1a-f.

The number of monthly declarants is more or less constant over time, or only slightly decreasing for the hospitality sector. Other accommodation is the smallest PPC with respect to this number, with around 250 monthly declarants, Restaurants is the largest one, with around 3000 monthly declarants. The number of quarterly declarants is strongly increasing for four of the six PPC's in this sector (Hotels, Restaurants, Fastfood, Catering). These numbers are by a factor of 5-16 larger than the number of monthly declarants (for the end of the considered period). The ratio of the turnovers is in the same order of magnitude.

3. Methods

In the previous section we introduced the published quarterly index series y_t^Q and the monthly auxiliary index series x_t. We define y_t^M as the 'true underlying' index series, which in practice would only be observed when all declarants are monthly declarants. The idea we discuss here is that although x_t is not a good proxy for a full monthly series, it might well be a good proxy for the monthly pattern within each quarter. This allows us to combine y_t^Q and x_t into a new estimate for the monthly index series y_t^M.

We write $\hat{y}_t^M$ for the monthly estimates, with the more extensive notation $\hat{y}_{t|M_x,M_y}^M$ which specifies that the estimate is based on the series x_t up to and including period M_x and the series y_t^Q up to and including period M_y. Furthermore we write $Q(t)$ for the last month of the quarter that t belongs to.

The common approach to compute $\hat{y}_t^M$ for periods t for which both the monthly and quarterly series are available, e.g. for $t \leq Q(T_Q) - 3$ is temporal disaggregation (TD). We discuss two methods for TD, i.e. the Chow – Lin (CL) method (Section 3.1.1) and the Denton (or Denton – Cholette) (DC) method (Section 3.1.2). TD is also possible with a structural time series model (STM), and especially a mixed frequency model, but this is not applied in this paper. We apply STM only for nowcasting.

When the quarterly figure for the most recent quarter is not available, e.g. when $t > M_y$, a nowcast method has to be applied for the last few months. These nowcasts can be computed using different nowcasting methods, which are discussed in Section 3.2 and 3.3.

In order to indicate which method of TD and which nowcasting method was used we write $\hat{y}_{t|M_x,M_y}^{M,\text{TD,nowcast}}$. In Section 3.4, alternative input series are discussed which could be used to estimate y_t^M. A discussion of the assumptions of the different nowcast methods is given in Section 3.5. Finally, in Section 3.6 we discuss how we will evaluate the different TD and nowcasting methods.

3.1 Temporal Disaggregation methods

3.1.1 Chow – Lin

There are several methods in the literature to translate the pattern of the auxiliary monthly series to the level of y_t^Q. The first one is a regression based technique (Chow and Lin, 1971, Fernandez, 1981). Chow and Lin (CL) assume a linear relation between the monthly auxiliary index series x_t, and the monthly (unobserved) variable of interest y_t^M. The CL method may also consider additional monthly series as auxiliary information, but here we describe it with the monthly auxiliary series x_t only. CL assumes the following Arima model for the monthly figures:

$$y_t^M = x_t\beta + u_t, \tag{4}$$
$$u_t = \rho u_{t-1} + \varepsilon_t, \tag{5}$$
$$\varepsilon_t \sim iid(0, \sigma^2). \tag{6}$$

The errors u_t are assumed to follow an AR(1)-process to account for autocorrelation in the series. The variable ε_t is white noise, with a variance that is constant over time. This set-up is very much a simplification of the true relation between the series y_t^M and x_t. However, it allows us to calculate the variance-covariance matrix of u_t, which is required in further estimation. Equations (4) - (6) are estimated at a quarterly frequency by using

$$y_t^Q = \mathrm{C}x_t\beta + \mathrm{C}u_t \quad \text{for } y_t^Q \neq \mathrm{NA} \tag{7}$$

And

$$C = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \ddots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 1 & 1 \end{bmatrix} \tag{8}$$

Since $\{u_t\}$ follows an AR(1)-process, the variance matrix $V = \mathrm{var}(u)$ can be estimated. Chow and Lin (1971, pp. 374-375) show that only ρ needs to be estimated and not σ^2. This allows them to use a relatively straightforward GLS-regression to estimate the ρ , u_t and β and therefore $\hat{y}_t^{M,\mathrm{CL}} = x_t\hat{\beta} + \hat{u}_t$.

3.1.2 Denton and Denton-Cholette

The other technique frames the disaggregation as a minimization problem (Denton, 1971, Dagum and Cholette, 2006). Without going into details we should mention that the original Denton method has some problems with the first few values of the series. The method aims at preserving all period-to-period movements of the auxiliary series, but additionally, it also minimizes the adjustment that is made to the first value of the monthly series. This has been fixed (see Dagum and Cholette, 2006, p.136, for an extensive description), so the general consensus amongst practitioners is to use the so called Denton-Cholette (DC) method instead of the Denton method.

In the Denton-Cholette model (the Denton model is a simplified version) the following entity is minimized over the monthly estimates $\hat{y}_t^{M,\mathrm{DC}}$:

$$Z = \sum_{t=2}^{T_m}\left(\frac{\hat{y}_t^{M,\mathrm{DC}}}{x_t} - \frac{\hat{y}_{t-1}^{M,\mathrm{DC}}}{x_{t-1}}\right)^2 \tag{9}$$

$$\text{s. t.} \quad y_t^Q = \frac{1}{3}\left(\hat{y}_{t-2}^{M,\mathrm{DC}} + \hat{y}_{t-1}^{M,\mathrm{DC}} + \hat{y}_t^{M,\mathrm{DC}}\right) \text{ for } t = 3,6,\dots,T_q. \tag{10}$$

The restriction (10) simply means that the three months in a quarter should be the quarterly index-value, which is also the case for CL. The idea behind the minimization (9) is to incorporate the movement of the indicator series, while avoiding sudden jumps that can arise because of level shifts in the quarterly series, that are absent in the indicator. It is easy to see that the optimum would be $\frac{\hat{y}_t^{M,\mathrm{D}}}{x_t} = \frac{\hat{y}_{t-1}^{M,\mathrm{D}}}{x_{t-1}}$ without the restriction of equation (10), but (10) ensures that the high and low frequent series are consistent within low frequent periods.

There are some other variants of this method. Equation (9) describes the proportional variant, there is also an additive one, which allows for non-positive values. We opt for the proportional version, because we don't consider negative index values.

The TD of both CL and DC is computed using R (R Core Team, 2018), using the R package forecast (Hyndman and Khandakar, 2008). This package is also used to compute the nowcasts as described in Section 3.2.

3.2 ARIMA nowcasting models

3.2.1 Simple extrapolation

The most straightforward way of nowcasting is to use CL or DC as described above and extend it with an extrapolation. With CL we can simply extrapolate (7) with x_t and the estimated $\hat{\beta}$ and $\hat{\rho}$, while for DC we simply let go of the restriction in (10) for the last month(s). Both will be referred to as (simple) extrapolation. Note that there is no ARIMA-model involved here. For CL, this means estimating the coefficients and residuals on the quarterly level up to T_Q and then applying those to all months up to T_M. For DC, it means that the nowcasted months have the exact same growth rate as x_t. If x_t has a high correlation with the quarterly series, both TD methods can give good results. However, we saw in the previous section that quite often the indicator has a (slightly) different pattern, mostly due to a difference in the trend. Therefore it might be better to separately forecast a quarterly and monthly trend.

When CL is applied, the method can be extended by adding a seasonal dummy as an additional auxiliary series in formula (4). This dummy is on the quarterly level. This method is referred to as "Simple extrapolation plus seas, CL".

3.2.2 BI-ratio

One way to account for the imperfections of the indicator is by using the Benchmark-to-Indicator-ratio (Bloem et al., 2001, also used by Daalmans, 2016). We explain the idea for CL, but for DC it works analogously. The basic assumption of this technique is that the relation between the indicator and the interpolated series, the β of Equation (4) is not constant over time. This implies we can improve the nowcast if we can predict this β_t. However, a time-varying β_t would not be identified properly in Equation (4), so instead for nowcasting month t with $t > M_y$ we compute the series $\hat{y}^M_{t|M_y,M_y}$ using TD. Then, for each month $t \leq M_y$ one can compute a ratio α_t by:

$$\alpha_t = \frac{\hat{y}^M_{t|M_y,M_y}}{x_t} \qquad (11).$$

This ratio is then forecasted using an ARIMA-model. Finally, we multiply the last months x_t, $t > M_y$, with the predicted BI-ratio to get our nowcast.

For automation purposes, we choose which ARIMA-model to use, with the Akaike Information Criterium (AIC, automated through the auto.arima function of R-package forecast, Hyndman and Khandakar, 2008). A disadvantage of the BI-ratio approach is that it can only use one indicator series. Therefore, in case additional series are used in the interpolation, it is more practical to use one of the other nowcast methods. This is not relevant for this research, as only one indicator per PPC is used, but for more general applications it should be noted.

3.2.3 Bridge

The last ARIMA based nowcasting method to discuss here is the so-called Bridge model (Baffigi et al., 2004). Here the order of disaggregating and forecasting is reversed, compared to the BI-ratio. The Bridge model first forecasts an additional quarter of y_t^Q with an AIC-selected ARIMA-model which may include additional regressors. This predicted quarter is added to the quarterly series, which serves as a sort of 'bridge'. Next, the observed monthly indicator is used to disaggregate this predicted quarter into three months.

There is one practical problem with interpolating the forecasted quarter. In the first and second month of the quarter that is nowcasted, the indicator series is only observed up to $T_q + 1$ and $T_q + 2$ (there is no problem for the third month). So the last month(s) of the indicator needs to be forecasted as well. In our setup, for this forecasting of the last months we also use an AIC-selected ARIMA-model as well.

We will test three versions of the Bridge model. The first version, referred to as 'Bridge model excl.', contains a simple quarterly forecast without additional regressors. The second version, referred to as 'Bridge model incl.', is extended by including the auxiliary monthly indicator x_t as a regressor, which is forecasted to have the series complete for the quarter. Then this monthly series has to be aggregated to the quarterly level (the mean of x_t within quarters). The third version, referred to as 'Bridge model incl plus seas.', a seasonal dummy is included as a second regressor. In this version, the seasonal dummy is also added in the TD-step, when CL is applied. In the case of DC, the seasonal dummy is added only when the quarterly series is forecasted.

3.3 Structural time series model

For this project, a special STM is developed to handle the different frequencies of the monthly and quarterly series. See Durbin and Koopman (2012) for a general introduction of structural time series models.

The series y_t^Q is extended with NA's (if necessary) to make sure that the series y_t^Q and x_t are equal in length.

The time series $\mathbf{y}_t = (\, y_t^Q, x_t)^{\,T}$ is modelled as:

$$\mathbf{y}_t = \mathbf{L}_t + \mathbf{S}_t + \mathbf{e}_t, \qquad (12)$$

with $\mathbf{L}_t = (\ell_t^y, L_t^x)^{\,T}$ the trend component, $\mathbf{S}_t = (s_t^y, S_t^x)^{\,T}$ the seasonal component and $\mathbf{e}_t = (e_t^y, e_t^x)^{\,T}$ the noise component. These three components are worked out as follows.

For the trend, we consider 4 different models.

The first STM model is shortly called local trend model without correlation. In this case, the trend of the monthly series is modelled as:

$$
\begin{aligned}
&L_t^x = L_{t-1}^x + \eta_t^x, \\
&E(\eta_t^x) = 0, \\
&cov(\eta_t^x, \eta_{t'}^x) = \begin{cases} \sigma_{L,x}^2 & \text{if} \quad t = t' \\ 0 & \text{if} \quad t \neq t' \end{cases}.
\end{aligned}
\tag{13}
$$

The trend of the quarterly series is modelled as

$$
\ell_t^y = \frac{L_{t-2}^y + L_{t-1}^y + L_t^y}{3},
\tag{14}
$$

where L_t^y is modelled with the local trend model, i.e.

$$
\begin{aligned}
&L_t^y = L_{t-1}^y + \eta_t^y, \\
&E(\eta_t^y) = 0, \\
&cov(\eta_t^y, \eta_{t'}^y) = \begin{cases} \sigma_{L,y}^2 & \text{if} \quad t = t' \\ 0 & \text{if} \quad t \neq t' \end{cases}.
\end{aligned}
\tag{15}
$$

And

$$
cov(\eta_t^y, \eta_{t'}^x) = 0 \text{ for all } t, t'.
\tag{16}
$$

The second STM model is shortly called local trend with correlation. This model is defined as the first STM model, but instead of eq. (16) we have:

$$
cov(\eta_t^y, \eta_{t'}^x) = \begin{cases} \varsigma_L & \text{if} \quad t = t' \\ 0 & \text{if} \quad t \neq t' \end{cases}
\tag{17}
$$

The third STM model is shortly called smooth trend without correlation. In this case, the trend of the monthly series is modelled as

$$
\begin{aligned}
&L_t^x = L_{t-1}^x + R_{t-1}^x, \\
&R_t^x = R_{t-1}^x + \eta_t^x, \\
&E(\eta_t^x) = 0,
\end{aligned}
\tag{18}
$$

$$
cov(\eta_t^x, \eta_{t'}^x) = \begin{cases} \sigma_{R,x}^2 & \text{if} \quad t = t' \\ 0 & \text{if} \quad t \neq t' \end{cases}.
$$

The trend ℓ_t^y of the quarterly series is modelled with (14), where L_t^y is modelled as smooth trend, i.e.

$$
\begin{aligned}
&L_t^y = L_{t-1}^y + R_{t-1}^y + \eta_t^y, \\
&R_t^y = R_{t-1}^y + \eta_t^y, \\
&E(\eta_t^y) = 0, \\
&cov(\eta_t^y, \eta_{t'}^y) = \begin{cases} \sigma_{R,y}^2 & \text{if} \quad t = t' \\ 0 & \text{if} \quad t \neq t' \end{cases}
\end{aligned}
\tag{19}
$$

and

$$
cov(\eta_t^y, \eta_{t'}^x) = 0 \text{ for all } t, t'.
\tag{20}
$$

The fourth STM model is called smooth trend with correlation. This model is defined with the equations of the third STM model, but replacing equation (20) by

$$cov(\eta_t^y, \eta_{t'}^x) = \begin{cases} \varsigma_R & \text{if} \quad t = t' \\ 0 & \text{if} \quad t \neq t' \end{cases}. \qquad (21)$$

Remark: Similarly as in (8), the trend of the quarterly series is the mean of the trend of the monthly series in three consecutive months, modelled with (14). This is only relevant for the third month of every quarter, as for the other months $y_t^Q = NA$. The model computes a trend L_t^y and ℓ_t^y for the first and second month of each quarter.

With the local trend model the value of the level at the end of the series determines the trend of the nowcasts, which will thus be constant for any nowcast horizon if the model does not allow for correlation between level disturbance terms. Otherwise the nowcast will be based partly on the evolution of the level of the monthly series. With the smooth trend model, the value of the slope at the end of the series determines the trend of all nowcasts, resulting in a linearly increasing or decreasing trend if the model does not allow for correlation between the slope disturbance terms. In the case of correlated slope disturbance terms, the nowcast will be based partly on the evolution of the observed smooth trend of the monthly series.It is left to further research whether an improvement of the nowcasts can be achieved with more complex (polynomial) trend models (Harvey, 1989).

The seasonal component of the monthly series is modelled with the so called trigonometric seasonal model (Durbin and Koopman, 2012) for monthly figures:

$$S_t^x = \sum_{l=1}^{6} S_{t,l}^x \qquad (22)$$

with

$$S_{t,l}^x = S_{t-1,l}^x \cos(h_l) + S_{t-1,l}^{x,*} \sin(h_l) + \omega_{t,l}^x,$$
$$S_{t,l}^{x,*} = S_{t-1,l}^{x,*} \cos(h_l) - S_{t-1,l}^x \sin(h_l) + \omega_{t,l}^{x,*},\ l = 1, \dots, 6$$
$$h_l = \frac{\pi l}{6},\ l = 1, \dots, 6$$
$$E(\omega_{t,l}^x) = E(\omega_{t,l}^{x,*}) = 0, \qquad (23)$$
$$\text{Cov}(\omega_{t,l}^x, \omega_{t',l'}^x) = \text{Cov}(\omega_{t,l}^{x,*}, \omega_{t',l'}^{x*}) = \begin{cases} \sigma_{\omega,x}^2 & \text{if} \quad t = t' \quad \text{and} \quad l = l' \\ 0 & \text{if} \quad t \neq t' \quad \text{or} \quad l \neq l' \end{cases}$$
$$\text{Cov}(\omega_{t,l}^x, \omega_{t',l'}^{x,*}) = 0 \text{ for all } l, l' \text{ and } t, t'.$$

The seasonal component of the quarterly series is modelled with an adapted version of the dummy seasonal model for quarterly series, which takes into account the missings.

$$s_t^y = \begin{cases} S_t^y & \text{if} \quad t \text{ third month of the quarter} \\ 0 & \text{if} \quad t \text{ first or second month of the quarter} \end{cases} \qquad (24)$$

with

$$S_t^y = \begin{cases} S_{t-10}^y & \text{if} \quad t \text{ first month of the quarter} \\ S_{t-1}^y & \text{if} \quad t \text{ second month of the quarter,} \\ -S_{t-3}^y - S_{t-6}^y - S_{t-9}^y + \omega_t^y & \text{if} \quad t \text{ third month of the quarter} \end{cases} \qquad (25)$$

$$E\left(\omega_t^y\right) = 0, \tag{26}$$
$$cov(\omega_t^y, \omega_{t'}^y) = \begin{cases} \sigma_{\omega,y}^2 & \text{if} \quad t = t' \\ 0 & \text{if} \quad t \neq t' \end{cases}.$$

In Van den Brakel and Krieg (2016), the trigonometric seasonal model (22)- (23) was applied for both the monthly and the quarterly series. There, the results were plausible, although there are only four observations available every year to estimate the monthly seasonal pattern. This does not apply here. It is found that only when the seasonal pattern is time independent, acceptable results can be achieved with the trigonometric seasonal model. Even then, the estimated seasonal pattern does not describe the differences between the months of a quarter well. The series considered in this paper have a time dependent seasonal pattern. Therefore, the trigonometric seasonal model (22), (23) is not an option, and (24)- (26) is developed.

With (24) - (26) only the quarterly seasonal pattern can be estimated. The third row of equation (25) is similar to the standard dummy seasonal model. In the third month of the quarter, new information becomes available. Therefore, the model allows for a (small) change via ω_t^y. It is assumed that the monthly seasonal pattern is constant within each quarter, which is achieved with the first and second row of (25).

Under the assumption that the seasonal patterns of the monthly declarants and the quarterly declarants is similar, it is desirable that this monthly pattern is adopted by the quarterly series. Nevertheless, (smaller) differences between the seasonal patterns of the monthly and the quarterly series should be taken into account. This cannot be achieved with the structural time series model. Instead, CL is applied, as will be described later.

The noise component is modelled with white noise. To take into account that both series are based partly on the same enterprises, two independent white noise variables are modelled:

$$E\left(\varepsilon_{t,j}\right) = 0,$$
$$cov\left(\varepsilon_{t,j}, \varepsilon_{t',j}\right) = \begin{cases} \sigma_{\varepsilon,j}^2 & \text{if} \quad t = t' \\ 0 & \text{if} \quad t \neq t' \end{cases} j = 1,2 \tag{27}$$
$$cov\left(\varepsilon_{t,1}, \varepsilon_{t',2}\right) = 0.$$

Then, $e_t^x = \varepsilon_{t,1} + \varepsilon_{t,2}$ and

$$e_t^y = \begin{cases} \dfrac{\varepsilon_{t-2,1} + \varepsilon_{t-1,1} + \varepsilon_{t,1}}{3} & \text{if} \quad t \text{ third month of the quarter} \\ 0 & \text{if} \quad t \text{ first or second month of the quarter} \end{cases}$$

In the current presentation, the sum of two independent white noise processes are used to model the measurement error in the monthly figures. The quarterly average of one of them is shared by the quarterly series. In this way the model accounts for the correlation in the measurement disturbance terms of the monthly and quarterly series, since the monthly and quarterly declarants are both used when the quarterly series is computed. Alternatively it is also possible to add the two independent white noise components to the quarterly series where one of them is shared by the monthly series. Since the process of the time series components are defined on a monthly frequency it is more straightforward to add the two white noise components to the monthly series.

The general way to proceed is to express the model in the so-called state-space representation and apply the Kalman filter to obtain optimal estimates for the state variables, see e.g. Durbin and Koopman (2012). Estimates for state variables for period t based on the information available up to and including period t are referred to as the filtered estimates. The filtered estimates of past state vectors can be updated, if new data become available. This procedure is referred to as smoothing and results in smoothed estimates that are based on the completely observed time series. It is also possible to predict values for time periods for which no data is available yet.

The analysis is conducted with software developed in OxMetrics in combination with the subroutines of SsfPack 3.0, see Doornik (2009) and Koopman e.a. (2008). All state variables are non-stationary with the exception of the errors. The non-stationary variables are initialised with a diffuse prior, i.e. the expectation of the initial states are equal to zero and the initial covariance matrix of the states is diagonal with large diagonal elements. The errors are stationary and therefore initialised with a proper prior. The initial values for the errors are equal to zero and the covariance matrix is available from the aforementioned model for the errors. In Ssfpack 3.0 an exact diffuse log-likelihood function is obtained with the procedure proposed by Koopman (1997).

The monthly estimates based on the quarterly series can be computed as $L_t^y + S_t^y + \varepsilon_{t,1}$. The noise term $\varepsilon_{t,1}$ is included here to make sure that the sum of the model estimates of three months are exactly equal to the input, which is the true value. However, since the model is used here to predict the monthly estimates, equality cannot be achieved, and the model prediction is $L_t^y + S_t^y$. The seasonal pattern of this model estimate is not computed on a monthly level, therefore CL is applied, see Section 3.1.1. To compute the seasonal monthly effect for month T_m, CL is applied to disaggregate the quarterly seasonal pattern $S_{t=3,}^y, S_{t=6,}^y, \dots, S_{t=Q(T_M),}^y$ using the estimates of the seasonal pattern of the monthly series $S_{t=1,}^x, S_{t=2,}^x, \dots, S_{t=Q(T_M)}^x$ as auxiliary series (see equation (22)- (23)).

Note that the last one or two values of these series are predictions, when T_m is the second or the first month of the quarter. Only when T_m is the last month of the quarter, no predictions are involved.

The result is called $\hat{S}_{t,}^y$ and the final estimate is computed as $\hat{y}_t^{STM} = L_t^y + \hat{S}_t^y$. The procedure results in estimates for the complete period, but only the estimate of the last month is used.

With this method, $\hat{y}_{t|M_x,M_y}^{STM}$ can be computed for all M_x, M_y by using the appropriate input series, extended with NA's if necessary.

The disaggregation can similarly be computed using DC (see Section 3.1.2). However, this is suboptimal because DC cannot handle negative values easily and they appear in the seasonal series. Nevertheless, we did investigate this option but we found that the predictions were generally worse than those obtained with CL, as can be seen in the results section, so here we won't further delve into DC disaggregation more deeply.

3.4 Alternative input series

3.4.1 Index series based on quarterly declarants only

An alternative approach, not further investigated in this paper, is to create a quarterly series based on enterprises that declare VAT on a quarterly basis and a monthly series based on enterprises that declare on a monthly basis. This approach was followed by Van den Brakel and Krieg (2016). In this approach both series are completely supplementary since enterprises contribute to one of the series only. There are, however, some practical disadvantages of this approach.

First, it would be necessary to select exactly the same enterprises which are used in the production of the quarterly series which are already published. Otherwise, it cannot be expected that the development of this series is reproduced. This is, at least, time consuming, as the decision of including or excluding enterprises is complex. Even if exactly the same enterprises are selected, there is no guarantee that the development of the published series is reproduced.

Second, dealing with enterprises which change from monthly to quarterly declaration (and vice versa) would become very important, as both series are influenced by them. The publication, i.e. increase or decrease, could depend on the choices which are made about them.

Third, when the quarterly series is disaggregated into monthly figures, this series has to be combined with the series based on the monthly declarants into one index of all declarants. For index series, this is not straightforward and can cause extra inaccuracy.

3.4.2 Modelling turnover series instead of index series

Another alternative approach is to model the series of the turnover instead of the index series. These series exist for both the monthly and the quarterly declarants, which means that four series have to be modelled, i.e. $\overleftarrow{z}_t^M$, $\vec{z}_t^M$, $\overleftarrow{z}_t^Q$, and $\vec{z}_t^Q$. This possibility is tested for a STM model. The results are compared with estimates based on index series, similar as the one described above (Section 3.3). Some small changes in the model are necessary. It was assumed before that it could be easier to model the sums of the turnovers instead of the index series. It is found that both models result in plausible estimates, where the estimates of the index series seem slightly more accurate. Details about the model for the four turnover series, the small changes in the index model and the estimates which are compared are described in Appendix 1. The results are not described in more detail. Based on this result, is was decided not to proceed with this approach.

3.5 Assumptions of nowcast methods

The nowcast methods described here are based on different assumptions about the quarterly and monthly series. The simple extrapolation assumes that the relationship between monthly and quarterly series is constant. Then the nowcast reflects the development of the monthly series. The BI-ratio-method allows for changes in this relationship, nevertheless is the nowcast strongly based on the development of the monthly series. With a bridge method, the nowcast is based on the development of the quarterly series. When the monthly series is included as auxiliary information (Bridge incl.) this information is also used, but only as far as there is a strong relationship. There are no strong and untested assumptions made about this relationship, instead it is derived based on the data.

The assumptions used in the STM are similar as with bridge. When an STM without correlation is applied, it is comparable with bridge excl, and STM with correlation is comparable with bridge incl. By STM, the choice for the local trend model or the smooth trend model determines how the prediction is computed. When the correlation is large, the two series develop similarly. Then the monthly series is used to predict the last quarter of the quarterly series. In the case of the local trend model, the trends L of both series increase or decrease together. In the case of the smooth trend model, the slopes R of both series increase or decrease together. This may result in a common increase or decrease of the trends, but it may for instance also result in a slowing growth in one series and increasing contraction in the other.

With bridge, this choice is determined by the ARIMA-model, which is chosen automatically.

Depending on which assumptions best fit the data, different nowcast methods can work best.

3.6 Evaluation of the methods

There are various ways to evaluate models. A common way is to compare out of sample predictions with actual outcomes. In case of nowcasting models this means pretending you are at a certain moment in the past and try to nowcast the next moment that by then is already known. In our case, in each month, we compute the nowcast $\hat{y}^M_{t|t,Q(t)-3}$. Unfortunately this nowcast cannot be compared with some real outcome, because the monthly pattern remains unknown. The best possible estimate for the monthly index is therefore based on one of the two TD methods using all available information that eventually becomes available. This is either CL or DC. Therefore we compare $\hat{y}^M_{t|t,Q(t)-3}$ with $\hat{y}^{M,\mathrm{CL}}_{t|T_Q,T_Q}$, the series we assume most closely mimics y^M_t for $t = 1,2,\dots,T_Q$, because it uses all the available information in a logical way. An alternative would be to compare $\hat{y}^M_{t|t,Q(t)-3}$ with $\hat{y}^M_{t|Q(t),Q(t)}$, which would resemble a monthly estimate available at the moment the last quarter is available. However, because both series are very similar, we consider the full series $\hat{y}^{M,\mathrm{CL}}_{t|T_Q,T_Q}$ to be our benchmark series. We found that the alternative estimates $\hat{y}^{M,\mathrm{DC}}_{t|T_Q,T_Q}$ based on DC are quite similar and could be used as benchmark as well.

For the six index series in the hospitality sector it was verified by sector experts, that these benchmark series, and especially their monthly patterns, are plausible. Another option of comparison is to compare $\hat{y}^{M}_{t|t,Q(t)-3}$ on the quarterly level, which has the advantage that y^{Q}_{t} is known.

It is important to note that in the production process, the most recent quarterly figure is available simultaneously with the most recent monthly figure. However, we think that for a comparison on the quarterly level it is better to make the comparison such that for the last three months nowcasts no quarterly data was available. Therefore we do not use this quarterly figure and compute the nowcasts $\hat{y}^{M}_{t|t,Q(t)-3}$ in all cases.
All the series that are nowcasted are available for the period January 2010 until June 2019. For the out of sample nowcasting we use the period July 2015 until June 2019, containing 48 months or 16 quarters.
A common way to indicate how one series mimics another is to calculate its mean absolute difference (MAD), which can be computed on the monthly and quarterly level. Furthermore, the production department also states that on the sector level (which may contain several PPC's) a MAD between nowcasts and the current published quarterly series should not be larger than 3 percentage points. The MAD is therefore our main evaluation criterion. The monthly MAD_m can be written as:

$$\mathrm{MAD}_m = \frac{\sum_{t=T_Q-47}^{T_Q}\left|\hat{y}^{M}_{t|t,Q(t)-3} - \hat{y}^{M,\mathrm{CL}}_{t|T_Q,T_Q}\right|}{48}, \tag{28}$$

and the quarterly MAD_q can be written as:

$$\mathrm{MAD}_q = \frac{\sum_{t=T_Q-47}^{T_Q}\left|\left(\sum_{t=Q(t)-2}^{Q(t)}\hat{y}^{M}_{t|t,Q(t)-3}\right)/3 - y^{Q}_{t}\right|}{16}. \tag{29}$$

The MAD can be computed on the PPC level by simply comparing the benchmark series with the nowcasted series. When series are compared on a more aggregated level, first these aggregated series are computed and then they are compared. So for instance the index series of the hospitality sector is a weighted average of our six PPC's, so we can construct a nowcast series for the hospitality sector by using the same weights. The weights are based on the VAT-share in the base year, similarly as used in the production of aggregated indices. The 3 percentage point criteria holds for the comparison of these two index series.

Additional to the weighted series, we also calculate the MAD for the unweighted series, which is simply the mean of the six series. The MAD for the weighted series evaluate how well indices at higher aggregation levels are nowcasted under the different models. The MAD for the unweighted series can be interpreted as an average evaluation of the different models for the PPC's where each PPC is equally important.

As mentioned before, we deviate slightly from the production process. In the production process the third month of a quarter can be computed simultaneously with the quarter. In our evaluation we ignore this fact and compare the average of three nowcasted months that are only based on monthly declarants with the published quarter. This means that the MAD_m should be smaller when the method is applied in production. On the other hand, there are some other reasons why the nowcasts do not match the final figures, like late response. This causes rather small differences between nowcast and final figures.

The MAD_m and MAD_q are two ways to evaluate and compare the different methods. In Appendix 2 two other measures are applied.

4. Results

Table 4.1 compares the different estimation methods for the monthly indices for the 6 PPC's of the hospitality sector using the monthly MAD, as defined in (28). Four types of series are compared; published quarterly series, disaggregated series using CL or DC, 11 different versions of ARIMA nowcasts (including the simple extrapolation) and 8 different versions of STM nowcasts. All these series are compared with the benchmarked series $\hat{y}^{M,\mathrm{CL}}_{t|T_Q,T_Q}$ as explained in Section 3.6, therefore the second row shows only zeros, because this concerns the benchmark series.

Type of series	Model\Hospitality sector	PPC1	PPC2	PPC3	PPC4	PPC5	PPC6	Unweighted mean	Weighted mean
Published	**Quarterly series**	6.3	13.9	4.6	6.2	8.5	4.3	4.8	4.1
Disaggregated	**Monthly series, CL**	0	0	0	0	0	0	0	0
	Monthly series, DC	0.3	2.7	0.1	0.4	0.5	0.6	0.5	0.2
ARIMA	**Simple extrapolation, CL**	3.1	19.3	2.6	7	5.6	5.3	4.3	2.4
	Simple extrapolation plus seas, CL	5.4	19	2.1	5.8	5.8	4.6	5.1	3.5
	Simple extrapolation, DC	2.3	20.4	2.3	7.5	4.4	3.4	3.9	2.1
	BI-ratio, CL	3.2	14	2.5	4.6	5.3	3.4	3.1	2
	BI-ratio, DC	2.4	18.5	2.2	6.6	4.3	2.9	3.6	2
	Bridge excl., CL	4.3	8.2	2.6	3.6	3.5	3	2.4	2
	Bridge excl., DC	4.2	8.7	2.6	3.5	3.5	3.2	2.5	2.1
	Bridge incl., CL	3.9	7.6	2.3	3.1	2.9	2.1	1.9	1.6
	Bridge incl., DC	3.9	8	2.3	3	2.9	2.2	2	1.7
	Bridge incl. plus seas, CL	9.3	18.5	6.3	7.1	10.9	6	6	5.1
	Bridge incl. plus seas, DC	9.7	18.8	6.2	7.2	10.9	6.4	6.3	5.3
STM	**Local trend model without cor., CL**	3.2	11.7	2.2	4.3	3.9	3	3.5	2.7
	Local trend model without cor., DC	4	25	3.3	7.3	3.7	5.4	5.3	3.6
	Local trend model with cor., CL	2.8	11.6	2	4.2	3.7	3	3.2	2.4
	Local trend model with cor., DC	4.1	25.5	3.3	7.3	3.5	5.2	5.1	3.2
	Smooth trend model without cor., CL	2.4	11.2	2	3.7	3.8	2.8	3	2.2
	Smooth trend model without cor., DC	3	25.2	3.6	7.4	3.5	5.6	4.6	2.9
	Smooth trend model with cor., CL	2.4	10.7	1.9	3.8	3.7	2.7	2.9	2.1
	Smooth trend model with cor., DC	3.3	24.8	3.5	7.3	3.4	5.5	4.4	2.8

Table 4.1: MAD_m scores of the six hospitality sectors. Grey: smallest three values for each column, PPC1: Hotels, PPC2: Other accommodation, PPC3: Restaurants, PPC4: Mobile food, PPC5: Catering, PPC6: Pubs

In the first row we also show the MAD_m of y_t^Q, which give us an indication of the potential scale of the MAD_m. Note that the MAD_m of y_t^Q is computed by assuming the value y_t^Q for the three months of each quarter.

The third row (MAD_m of $\hat{y}^{M,\mathrm{DC}}_{t|T_Q,T_Q}$) shows that TD with CL and DC give quite similar results.

We see that there are some substantial differences between the PPC's, not every PPC can be nowcasted with the same accuracy. In particular PPC2 has a high MAD_m. However, most PPC's have several models with MAD_m smaller than or close to 3 and this also holds for the weighted and unweighted average. And most important because it is a publication requirement, the column with the weighted mean index contains several values that meet this criteria.

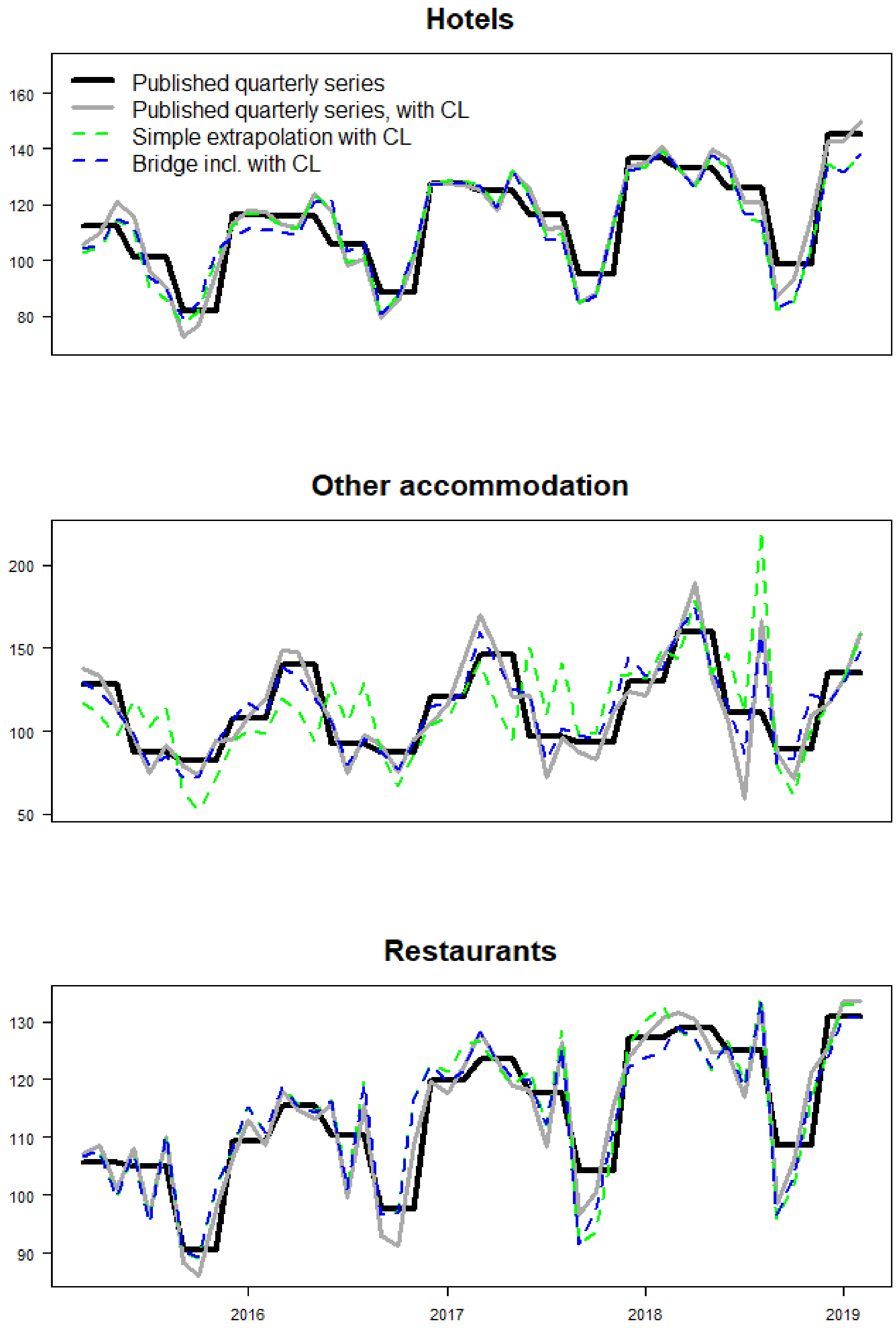
In each column the smallest three values are highlighted in grey. When we focus on these grey cells, a few things can be observed. First, in general 'Bridge incl., with CL' model does best, closely followed by the same model but with a DC disaggregation. Apparently it is worthwhile to nowcast the latest, unavailable quarter first by means of both its history (i.e. ARIMA) and the monthly index series x_t. Second, models that use a seasonal dummy never perform as the best. This may be because the dummies interfere with the seasonal pattern that is also present in x_t. Third, for some PPC's the STM models do best and in general the STM models (with CL) perform only slightly worse than the Bridge models while they do better than simple extrapolations or BI-ratio models.

Table 4.2 compares the different estimation methods for the quarterly indices for the 6 PPC's of the hospitality sector using the quarterly MAD, as defined in (29). Table 4.2 gives approximately the same picture as Table 4.1, especially when we consider the MAD_q of the weighted and unweighted mean series. Also on the quarterly level the Bridge incl. models perform best.

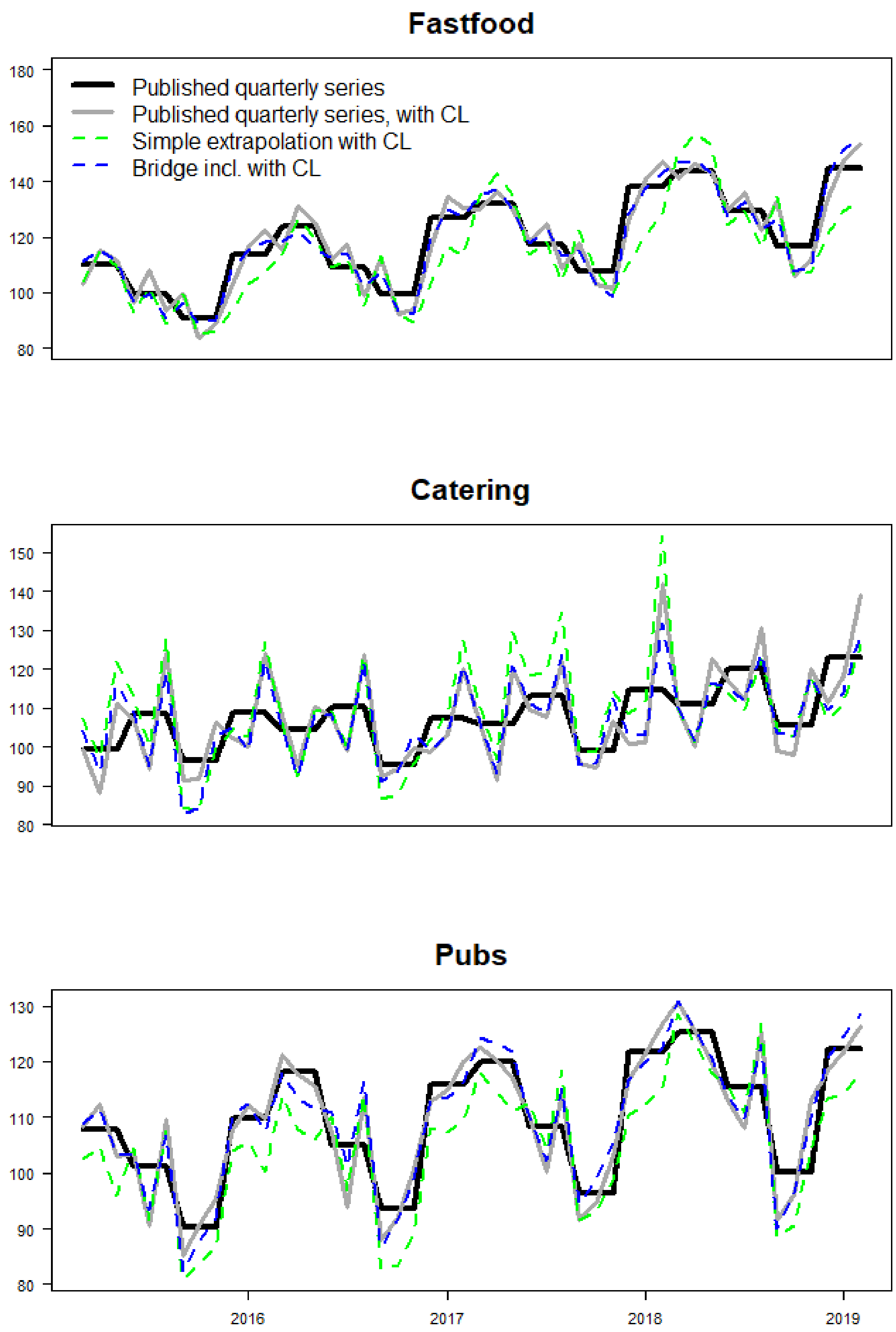
In order to see what the different MAD scores in Table 4.1 and 4.2 mean in practice, we look at the six graphs of the PPCs, in which we show the published series, the disaggregated series with CL, the simple CL extrapolation and the Bridge inc. with CL forecast. They are shown in graph 4.1a-c and 4.1d-f below.

Type of series	Model\Hospitality sector	PPC1	PPC2	PPC3	PPC4	PPC5	PPC6	Unweighted mean	Weighted mean
Published	**Quarterly series**	0	0	0	0	0	0	0	0
Disaggregated	**Monthly series, CL**	0	0	0	0	0	0	0	0
	Monthly series, DC	0	0	0	0	0	0	0	0
ARIMA	**Simple extrapolation, CL**	3.1	19.1	2.5	6.7	5.5	5.3	4.3	2.4
	Simple extrapolation plus seas, CL	3.8	12	1.9	3.6	3	3	3.4	2.5
	Simple extrapolation, DC	2.2	20.2	2.3	7.4	4.4	3.3	3.8	2
	BI-ratio, CL	3.2	12.2	2.4	4.6	5.1	3.4	2.9	1.9
	BI-ratio, DC	2.3	17.9	2	6.5	4.2	2.8	3.5	2
	Bridge excl., CL	3.6	5.2	2.2	2.4	2.2	2.5	1.5	1.3
	Bridge excl., DC	3.5	4.7	2.1	2.4	1.8	2.6	1.5	1.3
	Bridge incl., CL	3.8	4.8	2.2	2.1	2.2	1.8	1.1	1.2
	Bridge incl., DC	3.9	4.3	2.2	2.1	2	1.8	1	1.2
	Bridge incl. plus seas, CL	8.7	15.1	5.6	4.4	8.4	4.8	4.5	3.8
	Bridge incl. plus seas, DC	9.1	14.2	5.5	4.4	8.4	5.7	4.6	4
STM	**Local trend model without cor., CL**	3.1	10.2	1.8	3.7	1.8	2.7	3.2	2.5
	Local trend model without cor., DC	3.7	17.9	3	4.9	2.3	4.6	3.8	2.7
	Local trend model with cor., CL	2.6	10.2	1.7	3.7	1.7	2.7	2.9	2.1
	Local trend model with cor., DC	3.6	17.6	2.7	5.1	2.1	4.5	3.4	2.3
	Smooth trend model without cor., CL	2.2	9.3	1.5	3.3	2	2.5	2.8	2
	Smooth trend model without cor., DC	2.4	19	2.4	4.6	1.9	5	3.4	2.2
	Smooth trend model with cor., CL	2.1	9	1.5	3.4	1.9	2.4	2.7	1.9
	Smooth trend model with cor., DC	2.7	19.2	2.7	4.5	1.8	4.8	3.4	2.1

Table 4.2: MAD_q scores of the six hospitality sectors. Grey: smallest three values for each column, PPC1: Hotels, PPC2: Other accommodation, PPC3: Restaurants, PPC4: Fastfood, PPC5: Catering, PPC6: Pubs



Graph 4.1a-c: Published, disaggregated and nowcasted series of Hotels, Other accommodation and Restaurants.



Graph 4.1d-f: Published, disaggregated and nowcasted series of Fastfood, Catering and Pubs.

Graph 4.1a-f show how the published quarterly series (black line) is disaggregated by CL (grey line). Furthermore it can be seen that in general the blue dashed line (Bridge incl. with CL) follows the black line much more closely, albeit not always perfect. However, the blue dashed line is usually at least better than the green dashed line, which represents the more basic Simple extrapolation model. This illustrates the value of a better nowcasting model quite well.

In order to see if the conclusions from Tables 4.1 and 4.2 hold more general, we present similar results for six additional PPCs; Publishers, Legal activities, Accountants, Employment activities, Other Business Support, and Repair of household goods. Table 4.3 evaluates monthly nowcasts using the monthly MAD_m (28). Table 4.4 evaluates quarterly nowcasts using the quarterly MAD (29). For these six additional PPCs we get the same conclusion as for the hospitality PPCs in Table 4.1 and 4.2.

Type of series	Model\Hospitality sector	PPC7	PPC8	PPC9	PPC10	PPC11	PPC12	Unweighted mean	Weighted mean
Published	Quarterly series	7.9	7.3	7.4	3.9	7.1	4.9	5.5	5.4
Disaggregated	Monthly series, CL	0	0	0	0	0	0	0	0
	Monthly series, DC	1.4	0.4	0.8	3	0.5	0.5	0.5	0.5
ARIMA	Simple extrapolation, CL	8.7	5.8	9.5	5	5.4	5.2	2.9	3.4
	Simple extrapolation plus seas, CL	4.6	4.6	7.6	4.2	6.7	6	4	4.4
	Simple extrapolation, DC	12.1	5	10.1	5.3	7.2	6.2	3.4	4.1
	BI-ratio, CL	5.4	2.8	6	4.8	5.1	4.8	2.5	2.8
	BI-ratio, DC	9.3	4.3	6.6	5.9	6.2	5.7	2.9	3.5
	Bridge excl., CL	4.5	3.7	4.2	2.8	3.6	2.4	2.4	2.6
	Bridge excl., DC	5.4	3.6	4.5	4.9	4	2.2	2.6	2.9
	Bridge incl., CL	4.3	2.5	4.2	2.2	3.3	2.5	2	2.4
	Bridge incl., DC	5.2	2.3	4.5	4.2	3.4	2.4	2.2	2.6
	Bridge incl. plus seas, CL	13.3	9.8	8.6	4.9	7.8	6	6.6	6.7
	Bridge incl. plus seas, DC	12.3	10.1	9.4	7.8	8.2	6.5	7.1	7.2
STM	Local trend model without cor., CL	6	3.9	6.5	3.3	5.3	4.3	3.1	3.2
	Local trend model without cor., DC	7.5	8.1	11.7	3.4	9	7.9	4.4	5.2
	Local trend model with cor., CL	6.1	3.9	6.6	2.6	5.1	4	3	3.1
	Local trend model with cor., DC	7.3	8.2	12.1	2.9	8.7	8.1	4.4	5.2
	Smooth trend model without cor., CL	6.1	3.5	6.2	2.8	4.9	4.2	2.7	2.9
	Smooth trend model without cor., DC	7.3	7.8	11.3	3.4	9.8	7.9	4.2	5
	Smooth trend model with cor., CL	6.1	3.5	6.3	2.6	4.8	4	2.7	2.9
	Smooth trend model with cor., DC	7.4	7.8	11	3.9	9.7	7.9	4.2	4.9

Table 4.3: MAD_m scores and ranking scores of the six additional PPCs. PPC7: Publishers, PPC8: Legal activities, PPC9: Accountants, PPC10: Employment activities, PPC11: Other Business Support, PPC12: Repair of household goods

Type of series	Model\Hospitality sector	PPC7	PPC8	PPC9	PPC10	PPC11	PPC12	Unweighted mean	Weighted mean
Published	**Quarterly series**	0	0	0	0	0	0	0	0
Disaggregated	**Monthly series, CL**	0	0	0	0	0	0	0	0
	Monthly series, DC	0	0	0	0	0	0	0	0
ARIMA	**Simple extrapolation, CL**	8.5	5.5	9.3	5	5.2	5	2.9	3.4
	Simple extrapolation plus seas, CL	2.5	2.7	4.4	4	3.8	3.3	2	2.1
	Simple extrapolation, DC	11.9	4.9	10.1	4.9	7	6	3.3	4.1
	BI-ratio, CL	5.4	2.7	5.8	4.4	5	4.7	2.4	2.7
	BI-ratio, DC	9.3	4.2	6.1	4.3	6	5.6	2.8	3.4
	Bridge excl., CL	3.4	2.8	3	2.2	2.7	1.5	1.8	2
	Bridge excl., DC	2.7	2.9	3	3.2	2.6	1.4	1.8	2
	Bridge incl., CL	3.6	1.4	3	1.9	2.8	1.6	1.5	1.8
	Bridge incl., DC	2.8	1.6	3.1	2.4	2.8	1.6	1.6	1.8
	Bridge incl. plus seas, CL	12.7	7.8	5.4	3.9	4.7	3.5	5	4.8
	Bridge incl. plus seas, DC	11.1	8.5	6.4	6.7	4.8	4.6	5.7	5.5
STM	**Local trend model without cor., CL**	2.5	2.7	4.9	3.1	3.4	2.4	2	2.1
	Local trend model without cor., DC	5.6	5.7	7	2.9	6.7	4.1	3	3.4
	Local trend model with cor., CL	2.7	2.8	5	2.3	3	2.2	1.9	2.1
	Local trend model with cor., DC	5.5	5.6	7.4	2.3	6.5	4	3	3.4
	Smooth trend model without cor., CL	2.4	2.5	4.1	2.6	2.5	2.2	1.5	1.7
	Smooth trend model without cor., DC	5.4	5.7	6.4	2.4	7.4	3.8	2.5	2.9
	Smooth trend model with cor., CL	2.5	2.5	4.2	2.4	2.4	2.2	1.5	1.7
	Smooth trend model with cor., DC	5.4	5.7	6.8	2.4	7.3	3.8	2.7	3.1

Table 4.4: MAD_q scores and ranking scores of the six additional PPC's. PPC7: Publishers, PPC8: Legal activities, PPC9: Accountants, PPC10: Employment activities, PPC11: Other Business Support, PPC12: Repair of household goods

The results in the Tables 4.1 – 4.4 are all based on the MAD. In Appendix 2, two other evaluation methods are discussed. There, again, Bridge incl. with CL, is found to be a good choice.

Therefore we conclude that in nowcasting monthly index series of service PPC's by using monthly turnover data, in general the Bridge incl. with CL model seems to be a good choice.

5. Conclusion

In this paper different methods to compute monthly turnover figures are compared. There are two problems to be solved. First, for most enterprises only quarterly data is available. This makes temporal disaggregation necessary. Second, the quarterly data is only available after the monthly figures are computed. Therefore, nowcasting is needed. For the temporal disaggregation, two well-known methods (Chow and Lin, 1971, and Denton, 1971) are applied. These methods are combined with different nowcast methods. A general conclusion is that for most Primary Publication Cells (PPCs) there are a couple of models that are able to nowcast its series pretty well. For some PPC's it can be done more accurately than for others but on average the result is satisfying. The models that perform best seem to be the models that do not only rely on the monthly index series, like the ratio method, but also use the quarterly pattern to nowcast the most recent quarter. This property holds for both the so called Bridge models and STMs.

One method is found to perform best. This method computes a prediction of the present quarter, using an ARIMA-model and the monthly data as auxiliary information (a so called bridge model). This prediction is then used for temporal disaggregation using Chow Lin. A final estimate for monthly indices, if all monthly and quarterly information is available, is naturally based on temporal disaggregation with Chow Lin.

For the hospitality sector as a whole the nowcast deviates on average 1.6 index points from its benchmark value, which is based on temporal disaggregation with Chow Lin once all monthly and quarterly information is available. This difference is below the production requirement of a maximum of 3 percentage points, which implies it can be applied in practice. Some other methods including structural times series models also perform quite well, but we don't consider applying them in this setting in production, because the Bridge models perform slightly better and are much easier to apply.

In case there is a need for further improvement of the series three options are available. First, the use of additional regressors that are available until the last month. When there are regressors that are available for the most recent months and their pattern is related to the monthly index series of interest, it could be used to improve the nowcasts. The search for suitable regressors is outside the scope of this paper. A second option for improvement could be to inspect the data manually. Currently, the turnover data that underlies the quarterly index series are scrutinized manually. This implies that if there are say implausible high or low turnover values, this might be caused by administrative changes that are preferable excluded from the published indices. This could be because of a simple mistake (e.g. an extra zero), but also due to some accounting reality due to for instance a company that merge with another company or change from economic activity class. The data might then be improved manually. Such checks currently do not occur on the monthly level of turnover data, and so the quality of these data might be improved by introducing such a check. However, beside that this improvement is probably quite labour intensive, it may also be less valuable as the current process of manually improving the quarterly data. The reason is that the monthly nowcast relies on the (manually improved) quarterly index, while the quarterly index itself has no such series to rely on.

A third option for improvement might be to model the series not univariate as in this paper, but multivariate, i.e. using other PPC's as auxiliary information for the nowcast step. However, there is a risk of decreasing the quality of the nowcasts, as it might introduce spurious relations that cause noise in the nowcasts. More research is needed to find out which auxiliary information can be helpful.

Further research can also be done to include the largest enterprises for which Statistics Netherlands maintain primary data collection on a quarterly level (top-W). It is very costly for both Statistics Netherlands and the firms to change the survey-frequency to monthly. However, in some branches more than 95% of the turnover is determined by top-W, so it will be necessary, at least for some of the enterprises to collect data on a monthly frequency. It is important to investigate a way to incorporate these extra data.

In this paper, the methods are evaluated based on real data. A simulation could give more information about the accuracy of the methods. First, the true monthly pattern would be known in that case, and the accuracy of the disaggregation could be measured. Second, the development of the real series is quite regular. Generally, the trend is increasing in the entire period of 48 months. For some PPC's, it is decreasing for the entire period. In a simulation, the behavior of the methods close to turning points could be investigated. It is, however, not straightforward how simulation data could be produced. Therefore, this is left for further research.

References

Baffigi, A., Golinelli, R., and Parigi, G. (2004). Bridge models to forecast the euro area GDP. International Journal of Forecasting, 20(3), 447-460.

Baker, R., Blumberg, S. J., Brick, J. M., Couper, M. P., Courtright, M., Dennis, J. M., Dillman, D., Frankel, M. R., Garland, P. and Groves, R. M. (2010). Research synthesis AAPOR report on online panels. Public Opin. Q., 74(4), 711–781.

Banbura M., Giannone, D., and Reichlin, L. (2010). Nowcasting. CEPR Discussion Paper, DP7883. Available at SSRN: https://ssrn.com/abstract=1640405

Bloem, A.M., Dippelsman, R., and Maehle, N.O. (2001). Quarterly national accounts manual: Concepts, data sources, and compilation. Washington, D.C. International Monetary Fund.

Box, G. and Jenkins, G. (1970). Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day

Brakel, J. van den, and S. Krieg (2016), Estimating monthly turn- over with incomplete register data. Internal report. CBS, Heerlen.

Cholette, P.A. (1984). Adjusting sub-annual series to yearly benchmarks. Survey Methodology, 10, 35–49.

Chow, G. and Lin, A. (1971). Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series. The Review of Economics and Statistics 53, 372–375.

Daalmans, J.A., P. Vlag, S. Dek-Twaite, M. van der Geest, W. de Jong and C. Kaya (2015). Feasibility for Smart methods for monthly services estimates. Internal report. CBS, Den Haag.

Daalmans, J. (2016). Nowcasten van kwartaalomzetcijfers. Internal report. CBS, Den Haag.

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), 99-102.

Doornik, J.A. (2009). An Object-oriented Matrix Programming Language Ox 6. London: Timberlake Consultants Press.

Durbin, J., and Koopman, S.J. (2012). Time series analysis by state space methods, second edition. Oxford: Oxford University Press.

R. Fernández (1981), A Methodological Note on the Estimation of Time Series, The Review of Economics and Statistics, 63, 471-476

Harvey, A.C. (1989). Forecasting, structural time series models and the Kalman filter. Cambridge: Cambridge University Press.

Hyndman R.J., and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 26(3), 1–22. http://www.jstatsoft.org/article/view/v027i03

Koopman, S.J. (1997). Exact initial Kalman filtering and smoothing for non-stationary time series models. Journal of the American Statistical Association, 92, pp.1630-1638.

Koopman, S.J., N. Shephard, and J.A. Doornik (2008). SsfPack 3.0: Statistical algorithms for models in state space form. London: Timberlake Consultants Press.

Ouwehand, P. (2016). Nowcasten voor de horeca. Internal report. CBS, Den Haag.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Schouten, B. (2018). Snelle nowcasts met onvolledige respons, internal report. CBS, Den Haag.

Schreurs, G. (2012). Maandraming van de Commerciële dienstverlening als onderdeel van de Maandindicator Economische Groei. Internal report. CBS, Den Haag.

Appendix 1: Alternative structural time series models

The quarterly series which is published in production is about the index. The turnover sums $\overleftarrow{z}_t^Q$ and $\vec{z}_t^Q$, from which the index is computed (see formula (2)), are not saved. Therefore, series of turnover sums are computed for this project. Two series $\overleftarrow{z}_t^q$ and $\vec{z}_t^q$ are based on the quarterly declarants. This is different from formula (2), where the series are based on quarterly and monthly declarants. The other two series $\overleftarrow{z}_t^M$ and $\vec{z}_t^M$ are based on the monthly declarants.

From these series, the monthly and quarterly indexes are computed following formula (1), (2) and (3). The two index series are the input of a model which is similar as the model described in paragraph 3.3. The only difference is that $e_t^x = \varepsilon_{t,2}$ here. This change is necessary since the input series are based on different sets of enterprises. The model with smooth trends without correlation is used.

The four turnover series are also modelled in a different time series model. The series $\mathbf{Z}_t = (\vec{z}_t^q, \overleftarrow{z}_t^q, \vec{z}_t^M, \overleftarrow{z}_t^M)^T$ is modelled as

$$\mathbf{Z}_t = \mathbf{L}_t + \mathbf{S}_t + \mathbf{e}_t,$$

with $\mathbf{L}_t = (\vec{\ell}_t^y, \overleftarrow{\ell}_t^y, \vec{L}_t^x, \overleftarrow{L}_t^x)^T$, $\mathbf{S}_t = (\vec{s}_t^y, \overleftarrow{s}_t^y, \vec{S}_t^x, \overleftarrow{S}_t^x)^T$ and $\mathbf{e}_t = (\vec{e}_t^y, \overleftarrow{e}_t^y, \vec{e}_t^x, \overleftarrow{e}_t^x)^T$.

The trend components of the monthly series $\vec{L}_t^x$ and $\overleftarrow{L}_t^x$ are modelled as smooth trend, see (18). A covariance term of the disturbances of both series is assumed:

$$cov(\vec{\eta}_t^x, \overleftarrow{\eta}_t^x) = \begin{cases} \varsigma_{R,x} & \text{if} \quad t = t' \\ 0 & \text{if} \quad t \neq t' \end{cases}$$

For the trend components of the quarterly series, it is assumed that

$$\vec{\ell}_t^y = \vec{L}_t^y + \vec{L}_{t-1}^y + \vec{L}_{t-2}^y$$
$$\overleftarrow{\ell}_t^y = \overleftarrow{L}_t^y + \overleftarrow{L}_{t-1}^y + \overleftarrow{L}_{t-2}^y$$

These trend components $\vec{L}_t^y$ and $\overleftarrow{L}_t^y$ are modelled as smooth trend, see (18). A covariance term of the disturbances of both series is assumed:

$$cov(\vec{\eta}_t^y, \overleftarrow{\eta}_t^y) = \begin{cases} \varsigma_{R,y} & \text{if} \quad t = t' \\ 0 & \text{if} \quad t \neq t' \end{cases}$$

The covariances of the disturbances of monthly and quarterly series are assumed to be zero. This is similar as for the trend model, where the model without correlation is used.

Both monthly series are very similar, as they are based on more or less the same enterprises, except for the enterprises which are excluded due to changes. This means that a correlation close to 1 is expected. By modelling this correlation, the model estimates are probably more accurate. The same is true for the two quarterly series. It is also possible to model covariances between the disturbances of monthly and quarterly series. This is tested in the main part of the paper. Here we consider only one possibility.

The seasonal components of the monthly series $\vec{S}_t^x$ and $\overleftarrow{S}_t^x$ are modelled as trigonometric seasonal, see (22), (23).

The seasonal components of the quarterly series $\vec{s}_t^y$ and $\overleftarrow{s}_t^y$ are assumed to be equal $\vec{S}_t^y$ and $\overleftarrow{S}_t^y$ for the third month of a quarter, and 0 otherwise. $\vec{S}_t^y$ and $\overleftarrow{S}_t^y$ are modelled with the adapted dummy seasonal model (24) – (26).

$$\mathbf{e}_t = (\vec{e}_t^y, \overleftarrow{e}_t^y, \vec{e}_t^x, \overleftarrow{e}_t^x)^T$$

For the noise component, 4 white noise components $\varepsilon_{t,y,a}$, $\varepsilon_{t,y,b}$, $\varepsilon_{t,x,a}$, $\varepsilon_{t,x,b}$ are modelled, as in (27). Then $\vec{e}_t^y = \varepsilon_{t,y,a} + \varepsilon_{t,y,b}$, $\overleftarrow{e}_t^y$=$\varepsilon_{t,y,b}$, $\vec{e}_t^x = \varepsilon_{t,x,a} + \varepsilon_{t,x,b}$, $\overleftarrow{e}_t^x = \varepsilon_{t,x,b}$.

This way, it is taken into account that the two monthly series and the two quarterly series are based on almost the same enterprises.

The model is estimated using the Kalman filter, as described in Section 3.3. The model estimate for the quarterly series are now computed as

$$\widehat{\vec{z}_t^q} = \vec{L}_t^y + \frac{\vec{S}_t^y}{3}$$
$$\widehat{\overleftarrow{z}_t^q} = \overleftarrow{L}_t^y + \frac{\overleftarrow{S}_t^y}{3}$$

The aim of this appendix is to compare the accuracy of the model for the indexes and for the turnover sums. We focus on the results for the quarterly declarants without combining them with the monthly declarants. Furthermore, we restrict ourselves to estimates which have only a seasonal pattern on the quarterly level. The model for the turnovers estimates turnovers. These estimates can be used to compute an index, which could be compared with the estimates based on the model for the indexes. When the index is computed from model estimates of the turnover, the index depends on the decision in which period it should start with 100. Furthermore, a poor estimate for the turnover in one period influences all future estimates of the index. To avoid these problems, grow ratios are compared based on both models. The grow ratios are similar, and also similar to the published series (compared at quarterly level), and the grow ratios based on the model for the indexes are slightly closer to the published series than the ones based on the model for the turnover sums.

Appendix 2: Additional results

In Section 4 it is shown that Bridge incl., CL is a good method, based on the MAD_m and MAD_q. For most of the PPC's, this method has the smallest MAD_m and MAD_q, otherwise, it is not much worse than the best method. Besides these computations, also figures with times series of the estimates are evaluated, comparable to figure 4.1. These figures are not shown in this paper, but the general result is that Bridge incl., CL is one of the most accurate nowcasts for most of the periods, and never much worse than other methods.

Instead of showing these figures, we apply two other evaluation measures in order to show the good properties of Bridge incl., CL. In the first measure, it is counted how often in the 48 months the nowcast of another method is more accurate that the nowcast of Bridge incl., CL. When this happens in more than 50% of the periods, the other method can be considered more accurate than Bridge incl., CL. The results for the hospitality sector are shown in Table A1, in Table A2, the results for the 6 other PPC's follow. We see that percentages larger than 50% are quite rare in these tables.

In the second measure it is counted how often the prediction error, i.e. the absolute difference between the nowcast and ''true'' value, is larger than 8. The choice of the value of 8 is somewhat arbitrary. The idea is that, with an acceptable MAD_m of 3, an absolute difference of 8 can be considered an extremely bad nowcast. In Table A.3 and A.4, the percentages of these bad nowcasts are shown. We see that Bridge incl., CL has a relatively small percentage of such bad nowcasts.

		PPC1	PPC2	PPC3	PPC4	PPC5	PPC6	Unweighted mean	Weighted mean
ARIMA	Simple extrapolation, CL	17	31	40	29	17	17	17	17
	Simple extrapolation plus seas, CL	21	27	60	25	25	21	21	23
	Simple extrapolation, DC	33	25	48	29	29	33	21	27
	BI-ratio, CL	31	19	50	38	29	31	27	35
	BI-ratio, DC	29	23	56	27	33	29	17	27
	Bridge excl., CL	25	35	50	23	42	25	25	19
	Bridge excl., DC	27	46	46	29	38	27	29	23
	Bridge incl., CL	0	0	0	0	0	0	0	0
	Bridge incl., DC	52	50	40	56	48	52	52	54
	Bridge incl. plus seas, CL	21	23	25	21	13	21	15	13
	Bridge incl. plus seas, DC	19	29	25	19	19	19	13	8
STM	Local trend model without cor., CL	44	35	58	38	33	44	27	21
	Local trend model without cor., DC	29	25	40	31	42	29	19	19
	Local trend model with cor., CL	46	38	63	46	35	46	27	21
	Local trend model with cor., DC	25	21	42	33	40	25	23	25
	Smooth trend model without cor., CL	33	38	52	44	35	33	29	29
	Smooth trend model without cor., DC	33	29	44	27	48	33	27	35
	Smooth trend model with cor., CL	35	40	54	44	38	35	31	27
	Smooth trend model with cor., DC	35	25	40	27	50	35	25	38

Table A1: results percentage better than brigde incl CL, horeca, PPC1: Hotels, PPC2: Other accommodation, PPC3: Restaurants, PPC4: Fastfood, PPC5: Catering, PPC6: Pubs

		PPC7	PPC8	PPC9	PPC10	PPC11	PPC12
ARIMA	Simple extrapolation, CL	23	29	27	21	33	27
	Simple extrapolation plus seas, CL	38	31	33	17	25	25
	Simple extrapolation, DC	21	33	23	31	31	21
	BI-ratio, CL	46	48	35	25	31	48
	BI-ratio, DC	19	38	33	19	27	29
	Bridge excl., CL	50	29	54	40	46	58
	Bridge excl., DC	38	31	48	27	40	54
	Bridge incl., CL	0	0	0	0	0	0
	Bridge incl., DC	44	63	48	25	46	52
	Bridge incl. plus seas, CL	25	21	27	31	25	29
	Bridge incl. plus seas, DC	25	17	31	23	23	27
STM	Local trend model without cor., CL	50	35	33	38	35	33
	Local trend model without cor., DC	25	23	25	35	33	15
	Local trend model with cor., CL	46	35	31	40	31	35
	Local trend model with cor., DC	33	25	23	48	31	13
	Smooth trend model without cor., CL	42	33	29	40	33	29
	Smooth trend model without cor., DC	29	27	25	29	40	17
	Smooth trend model with cor., CL	42	33	29	44	35	35
	Smooth trend model with cor., DC	29	27	27	31	35	19

Table A2: results percentage better than brigde incl CL, other PPC's, PPC7: Publishers, PPC8: Legal activities, PPC9: Accountants, PPC10: Employment activities, PPC11: Other Business Support, PPC12: Repair of household goods

		PPC1	PPC2	PPC3	PPC4	PPC5	PPC6	Unweighted mean	Weighted mean
ARIMA	Simple extrapolation, CL	23	77	0	31	23	23	33	15
	Simple extrapolation plus seas, CL	13	65	2	25	27	13	33	19
	Simple extrapolation, DC	2	75	0	35	17	2	29	6
	BI-ratio, CL	0	75	2	17	23	0	17	4
	BI-ratio, DC	2	73	2	33	17	2	17	4
	Bridge excl., CL	2	48	0	8	6	2	10	6
	Bridge excl., DC	4	44	0	8	6	4	8	2
	Bridge incl., CL	0	42	0	10	6	0	0	0
	Bridge incl., DC	0	46	0	8	0	0	2	0
	Bridge incl. plus seas, CL	31	71	29	38	63	31	48	38
	Bridge incl. plus seas, DC	33	71	29	38	63	33	52	40
STM	Local trend model without cor., CL	4	58	2	17	13	4	17	10
	Local trend model without cor., DC	19	71	10	35	6	19	35	23
	Local trend model with cor., CL	4	54	0	15	8	4	15	6
	Local trend model with cor., DC	19	71	6	35	6	19	33	21
	Smooth trend model without cor., CL	0	46	0	6	8	0	17	4
	Smooth trend model without cor., DC	17	63	8	25	8	17	33	21
	Smooth trend model with cor., CL	0	46	0	6	8	0	15	4
	Smooth trend model with cor., DC	17	67	10	27	8	17	29	19

Table A3: results percentage worse than 8 (for 6 PPC's) or worse than 6 (for means), hospitality sector, PPC1: Hotels, PPC2: Other accommodation, PPC3: Restaurants, PPC4: Fastfood, PPC5: Catering, PPC6: Pubs

		PPC7	PPC8	PPC9	PPC10	PPC11	PPC12
ARIMA	Simple extrapolation, CL	35	31	48	15	25	21
	Simple extrapolation plus seas, CL	21	10	38	0	38	31
	Simple extrapolation, DC	56	19	54	27	42	27
	BI-ratio, CL	27	2	21	17	21	21
	BI-ratio, DC	48	15	23	31	27	23
	Bridge excl., CL	17	8	10	0	10	4
	Bridge excl., DC	27	8	10	19	17	2
	Bridge incl., CL	15	0	10	2	2	2
	Bridge incl., DC	23	0	10	10	4	4
	Bridge incl. plus seas, CL	60	42	46	17	38	27
	Bridge incl. plus seas, DC	58	42	48	29	42	31
STM	Local trend model without cor., CL	29	8	29	4	27	17
	Local trend model without cor., DC	50	25	52	6	38	48
	Local trend model with cor., CL	27	8	29	2	23	15
	Local trend model with cor., DC	52	25	54	4	35	48
	Smooth trend model without cor., CL	29	10	27	4	19	19
	Smooth trend model without cor., DC	48	19	50	8	38	52
	Smooth trend model with cor., CL	27	10	27	2	19	17
	Smooth trend model with cor., DC	48	19	46	13	35	52

Table A4: results percentage worse than 8, other PPC's, PPC7: Publishers, PPC8: Legal activities, PPC9: Accountants, PPC10: Employment activities, PPC11: Other Business Support, PPC12: Repair of household goods

Explanation of symbols

Empty cell	Figure not applicable
.	Figure is unknown, insufficiently reliable or confidential
*	Provisional figure
**	Revised provisional figure
2017–2018	2017 to 2018 inclusive
2017/2018	Average for 2017 to 2018 inclusive
2017/’18	Crop year, financial year, school year, etc., beginning in 2017 and ending in 2018
2013/’14–2017/’18	Crop year, financial year, etc., 2015/’16 to 2017/’18 inclusive

Due to rounding, some totals may not correspond to the sum of the separate figures.

Colophon

Publisher
Centraal Bureau voor de Statistiek
Henri Faasdreef 312, 2492 JP Den Haag
www.cbs.nl

Prepress
Statistics Netherlands, CCN Creation and visualisation

Design
Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contactform: www.cbsl.nl/information