Discussion paper

Productivity in the Dutch Public Sector

The case of libraries and fire services

Enrico Maroni

June 2021

Content

Summary

Efforts to measure productivity in the public sector have been revitalized in recent decades, as demand for better government spending accountability and policy evaluation strengthened. Productivity can easily be defined as a measure of the amount of output obtained by a given set of inputs; however, carrying out this calculation for public organizations is a complex and long-discussed challenge. That is, computing total factor productivity in public institutions requires a higher number of measurement choices than for private ones. As it will be tackled below, this is mainly due to lack of meaningful price information, which renders output direct observation the only feasible methodology. Nevertheless, explaining output is not the only problem, as quality adjustments and aggregation issues also require careful consideration.

This study proposes a methodological discussion of these challenges, which are then applied in the calculation of productivity in two Dutch public institutions: libraries and fire services. For the latter a novel monetary-based quality adjustment is proposed to account for real estate and people's health damages. It is found that TFP sharply increased for libraries, while evidence is less clear for the fire department. The magnitude of these changes is a rather volatile matter due to missing data, which forces the use of sensitivity analysis as a primary analytical tool. To this extent, it will be highlighted that the availability of more detailed output and cost structure information are desirable future implementations.

Keywords

Public sector, productivity, total factor productivity

1. Introduction

When it comes to economics, gross domestic product (GDP) is without any doubt the most well-known and popular indicator. Given its importance, one would expect GDP calculations to be based on rigorous and scientific measures of the economy, but this is not entirely the case. Diane Coyle (2015) defines GDP as a "made-up entity", which aggregates all goods and services produced by the economy with extremely complex statistical techniques and, in some cases, arbitrary assumptions. The latter are particularly crucial to measure the output of some sectors of the economy, such as the financial industry (Basu, Inklaar and Wang, 2011), and, as it will be explained on these pages, the public sector.

The theme of correctly measuring the output of the public sector is not only crucial for the computation of GDP but also for considerations about its productivity levels. In fact, when output measures for public institutions are not available, they are accounted for according to the so-called output=inputs method, which makes productivity always equal to one. Yet, as explained below, learning about productivity in public bodies is important not only for policy evaluation purposes but also for the economy as a whole, since the public sector represents a large and growing area, especially in developed countries such as The Netherlands.

The main objective of this paper is to propose a comprehensive analysis of the main issues one needs to consider when trying to overcome the output=inputs method to calculate the development of total factor productivity (TFP) in public bodies. That is, when it comes to public institutions, the lack of meaningful price information (public services and products are often offered at non-market prices or even free of charge) makes it necessary to directly construct volume output indices by carefully assessing the institution’s production process.

This initial discussion and the idea of aggregating different output indices using administrative cost information build on the Atkinson Report (Atkinson, 2005), which laid the foundation for output and productivity measurement in the public sector. Secondly, as discussed in further detail below, Schreyer (2012) shows the importance of considering outcome to determine quality adjustment of output indices in health and education services. Even though this paper focuses on different parts of the public sector, the methodological framework proposed by Schreyer is crucial to highlight the importance of differentiating products and directly including quality changes in the output volume measure. Thirdly, once quality-adjusted indices are calculated and ready to use, references to Diewert (2017) are made about how to correctly combine input and output indices, taking into consideration the shape of the potential production function to estimate TFP changes.

Nevertheless, the discussion does not stop at a theoretical level, but it goes further and tackles the implementation of two practical analyses for the Dutch libraries and fire brigades. These case studies allow to demonstrate the importance of initial measurement choices, on which final productivity levels heavily depend. In particular, the final output of public libraries is based on the combination of loans, visits and the number of activities organized, the quality adjustments taking physical and digital collection variety into account. Although the aggregation of these dimensions relies on a sensitivity analysis, as information about consumers' preferences or costs is not available, the numbers safely demonstrate that productivity increased at very high annual rates, ranging between 6% and 10%, in the years 2015-2018. The analysis of the fire department, in turn, divides the production process in two main parts: (i) firefighting interventions and (ii) prevention activities. The calculation focuses on the former dimension (as data about the latter one are not available), correcting the number of interventions by the average annual damage caused by fires. That is, this measure gauges the effectiveness of firefighters in intervening on the fire scene, as damages are a direct function of their preparation and organization like skills, response time, abilities, etc.

Accounting for fire damages is not a novel approach (Bouckaert, 1992; Jaldell, 2002); however, this paper explores the use of a monetary-based aggregation of (i) real estate damages, (ii) victims, and (iii) injured people, avoiding aggregation arbitrariness and trade-offs. The resulting TFP figures are negative (-5% between 2013 and 2019) because of the massive drop in the number of interventions, which is a consequence of fewer fires. Therefore, it may be that this reduction is caused by more (and better) prevention activities, which are the second task of the department. Although actual data are not available, a projection accounting for this second aspect is proposed, showing that TFP may turn positive.

In conclusion, this paper shows that measuring productivity in the public sector requires to carefully consider a variety of measurement choices, as they play a bigger role in determining final results than in the 'private market' context. An example of this relates to how volunteers' contribution is measured and accounted for: as shown, if they are paid below their marginal productivity, the actual productivity may be considerably lower than the original calculation. Following the crucial role of measurement assumptions, the importance of improving the data collection process is highlighted at the end of the analysis. Among others, information about the cost structure is particularly important to aggregate outputs. The structure of the paper is as follows: section 2 discusses the theoretical framework, section 3 presents the data and the two case studies, while section 4 draws some conclusions and recommendations. In addition, a brief discussion about National Accounts and inputs' quality adjustments is presented in the appendix.

2. Background and productivity

2.1 Why productivity

2.1.1 The basic concepts of productivity

Syverson's comprehensive survey on productivity and its relative determining factors (Syverson, 2011) simply defines it as the measure of the quantity of output obtained by a given set of inputs. In other words, productivity is simply the output-input quantities ratio:

$$Productivity = \frac{Outputs}{Inputs} \tag{1}$$

As for GDP growth, when assessing productivity, the focus is also on quantities discarding any effects caused by changes in prices, which do not say anything about how well inputs are transformed into outputs.

Even if the definition of productivity seems pretty straightforward, Diewert and Nakamura (2005) highlight that it represents a nuanced concept, which needs some further specifications. First of all, productivity is a broad term that encompasses different types of measures. Firstly, single-factor productivity (SFP) represents the comparison between the quantity of output produced with the quantity of only one of the inputs used in the production process. In the particular case the input analysed is labour, SFP becomes labour productivity (LP), which measures how much output is produced by a unit of labour (usually, hours worked). Secondly, when all the inputs are included in the computation of productivity, this index is called total factor productivity (TFP)1. It is worth emphasizing that single-factor estimations do not consider the intensity of use of the other inputs. Therefore, if the final scope is to have a clear overall picture about how efficiently factors are combined in the production process, we should aim to determine TFP. Moreover, attention should be given to the difference between the value of productivity change ΔTFP and the productivity level (TFP). If productivity growth is higher for a firm than for another, this does not imply that the former is more productive than the latter (Diewert and Nakamura, 2005).

Furthermore, often TFP is interpreted as a proxy of technical development. However, Lipsey and Carlaw (2000) argue that changes in TFP are different from changes in technology. The same view is sustained by Hulten (2010), who highlights that different factors may cause technological shifts (technical innovations, organizational and institutional changes, societal attitudes, etc.). Moreover, Basu and Fernald (2002) show the existence of meaningful gaps

[1] It is worth emphasising that sometimes TFP is referred to as MFP (Multi-factor Productivity). In practice, the difference between the two concepts is not very clear, as in both cases the main objective is to include as many inputs as possible.

between productivity and technology (mainly due to frictions in output and factor market), the latter one being less correlated with output than the former.

2.1.2 Why are productivity and its measurement important?

Although productivity is an abstract and constructed concept, economists care a great deal about its improvements and changes overtime, as it closely correlates with the overall economic growth and per-capita income growth. Therefore, there is a vast literature that considers productivity as a crucial explanatory factor, both for macro- and firm-level economic models.

For the first set of studies, according to the so-called development accounting models[2], productivity is considered to play a pivotal role in accounting for income differences across countries. There is a consensus among researchers that human and physical capital account for 30-50% of country differences, the residual being explained by TFP variations (Hsieh and Klenow, 2010). Furthermore, Englander and Gurney (1994) explain how productivity growth is decisive for countries to improve their real incomes and welfare.

On a more firm-based level, new trade and foreign direct investment (FDI) models (Melitz, 2003; Helpman, Melitz and Yeaple, 2004) hinge on productivity differences across companies to introduce the concept of firm heterogeneity. The main idea behind these models is that productivity differences can explain which companies are to develop foreign activities, engage in FDI, sell domestically, or exit the market. That is, TFP levels determine if a firm will perform international activities and thrive or will leave the market in the face of competition.

Furthermore, the relevance of accurately measuring productivity is especially clear if analysed under a dynamic perspective. Stiglitz, Sen and Fitoussi (2009, page 7), in their study for the French Government, emphasize the importance of getting measurements (more) right, writing: "*what we measure affects what we do; and if our measurements are flawed, decisions may be distorted*". In other words, since future policy design and management decisions are based on the analysis of current data, the first necessary step is having right and steady measurement frameworks to rely on.

2.1.3 Why is it even more important for the Government sector?

Until now, we have discussed productivity and its importance under a general point of view, which has not considered particular sectors or markets. However, the objective of this paper refers to productivity in the Government and public organizations, so that some specific features need to be highlighted.

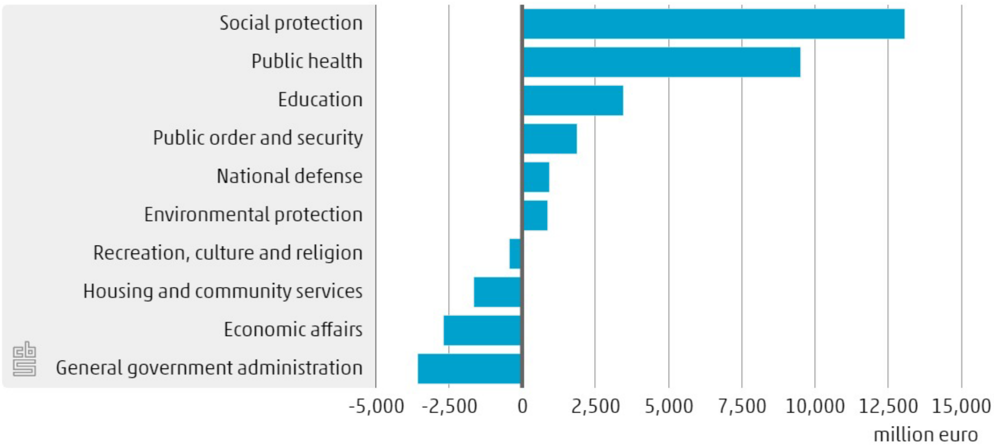
Firstly, governmental organizations are funded with resources coming from the taxpayers, who are highly interested to know how their money is being spent (Gruen, 2012). Moreover, there is an increasing concern about the quality of public

[2] For a thorough analysis of developing accounting frameworks, the reader should refer to the survey carried out by Francesco Caselli (2005).

finances, which can directly affect the country's performances on international markets, influencing investors' behaviour (Atkinson, 2005).

Secondly, implementing a measure of productivity in public organizations is the first and necessary step for a thorough system of public policy evaluation. According to the (Australian) Productivity Commission (2017), it is pivotal to understand which the drivers of efficiency are in order for the policymaker to improve the quality of new legislations. In the Netherlands, this objective is pursued by the project Operatie Inzicht In Kwaliteit (OIIK), implemented by the Ministerie van Financiën (Ministry of Finance). The operation aims at improving the social impact of policies, making them understandable, feasible and positively impactful for citizens and companies. Certainly, this is not possible if a general assessment framework is not in place. Moreover, the extent to which productivity proxies the degree of efficiency in the productive process can also play a role in determining funding and setting budgets. Data about the Dutch Government's spending in different public sectors (COFOG classification), displayed in figure 2.1.3.1, show that policy interventions can cause large financial shifts, as happened between 2010 and 2018. Consequently, it would be important to couple these spending figures with their relative impact on productivity, to understand whether increases in funding have led to higher output levels rather than a lower TFP, laying the groundwork for a continuous monitoring and evaluation system.

Figure 2.1.3.1. Shift of Government spending between sectors 2010-2018



Thirdly, the government sector is strongly interconnected with all the other sectors of the market. Consequently, its developments can trigger spill overs in other parts of the economy (e.g. high-quality education system increases students' skills and, in turn, their future performances on the job market). This is particularly noteworthy if we account for the productivity slow-down Western countries have been witnessing since the 2008 crisis. Fernald and Inklaar (2020) show that TFP growth rates were positive in almost every developed country during the period 1995-2007, while being close to zero from 2007 to 2015. They highlight how the figure is even more severe for Europe: the Northern countries[3] are keeping pace with the U.S. TFP level but are not converging anymore, while the Southern

[3] Belgium, Finland, France, Germany, Netherlands, and the UK.

countries[4] are actually diverging. For this reason, correctly measuring productivity in the Public sector, which sums up to a significant part of the total economy (Dutch Government's spending was 42% of GDP in 2019[5]), would play a big role in defining actual national TFPs. Furthermore, as shown by Basu, Pascali, Schiantarelli and Serven (2012), countries' welfare levels are summarized, to a first order, by productivity (TFP). That is, information about total factor productivity (coupled with the level of national capital stock) can be used to explain across and within countries (households) welfare differences, once again, reflecting the importance of accurately measuring TFP.

Finally, Lau, Lonti, and Schultz (2017) highlight the importance of considering a dynamic perspective. In fact, there are some important trends that need to be accounted for: (i) ageing populations will trigger higher demand for public services (e.g. health), (ii) decreasing number of hours worked per person require better labour productivity to maintain the same output level, (iii) declining trust in Governments requires politicians to improve the quality of the services offered in order to change people's perception.

2.2 Terminology and non-market peculiarities

Hitherto, the main purpose of the discussion has been to give some background knowledge and to introduce the general concept of productivity. From now on, productivity in governmental organizations will be the focus. Nevertheless, before tackling the methods (and the issues) we can employ to measure productivity, it is worth defining some terminology and highlighting some peculiarities of the non-market sector.

First, we have already mentioned the concept of input and output; however, if we need to formally define them:

- Inputs are the goods and services employed in the productive process. They usually comprise capital services, labour and intermediate consumption.
- Outputs are suitably differentiated and are the number of constant-quality actions or activities (in the case of services), and the number of constant-quality physical units (in the case of goods) (Schreyer, 2012).

Second, output and outcome are two completely different concepts. Outcome is a state valued by consumers, which is influenced by many factors, and one of them may be the level of output. For example, in the case of the health sector, outcome is the population's health state, while output is the number of activities carried out by hospitals (e.g. surgeries). The bottom line is that outcome is not only influenced by output but also by other variables (e.g. lifestyle, sport, etc.). The difference between the two concepts was already highlighted by Hill (1977), who defined outcome as the purpose for which goods and services are produced, while outputs, as the goods and services themselves. Although output and outcome are

4 Spain and Italy.
5 https://opendata.cbs.nl/statline/#/CBS/nl/dataset/84114NED/table?ts=1608842616288

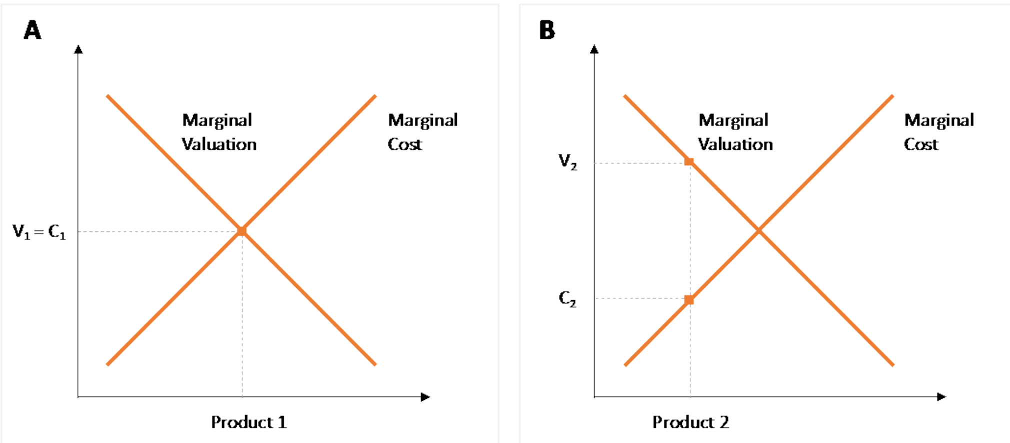
two different measures, in practice they are interconnected and, to fully understand and measure output, some reference to outcome should be employed (Schreyer, 2012). For example, if a certain type of surgery (an output) is improved to lead to a lower mortality level (an outcome), then we should consider that the quality of this output has improved.

Next, what are the most important differences between market and the Public sector? Prices are used as weights to sum up volume measures of different products in market contexts. This is possible because, under competitive conditions, prices (p_i) are believed to represent both user's marginal valuation and producer's marginal cost, so that total value of the transactions considered can be decomposed as follows:

$$V = \sum_{i=1}^{n} (x_i * p_i) = \sum_{i=1}^{n} (x_i * v_i) = \sum_{i=1}^{n} (x_i * c_i) \quad (2)$$

Where x_i represents the quantity of each product, v_i the marginal valuation by the consumers and c_i the marginal cost of supplying. Firms are believed to supply products until the point where marginal costs and prices are equalized, or more generally, at the point where prices will equal average costs. Schreyer (2012) explains this concept combining the producer.

Figure 2.2.1. Marginal valuation and marginal cost in market and non-market contexts



However, when it comes to the non-market sector, specifically Government production, this mechanism does not necessarily hold. The absence of relevant economic transactions renders it difficult to identify prices and quantities.

The first big problem is that public services (and goods) are usually provided at non-meaningful prices (sometimes for free), which, consequently, are not representative of neither users' marginal valuation nor producers' marginal cost (as depicted in figure 2.2.1, panel B). The second big set of problems relates to the difficulty of identifying correct quantity measures. This issue will be tackled later, when discussing how to estimate output for public institutions. Here, it is worth noting that this problem is more severe for collective services (e.g. police, defence, environmental protection, etc.), while for individual services (e.g. education, health, etc.), some direct measures are already in place in many countries. However, Lau, Lonti and Shultz (2017) highlight that a significant part of OECD countries' public spending is made up by collective services, with an average value close to 30% of the overall public expenditures in those countries.

2.3 Outputs in the Public Sector

2.3.1 Output=Input and the problems with volume

According to the Atkinson Review (2005), which represents a well-recognized attempt to lay the groundwork for productivity measurements in governmental organizations, a great majority of countries, from the 1960s until the end of the 1990s, have measured output produced by public bodies as a value equal to the sum of the inputs used in the production process. This practice is referred to as the output=input convention, and it has been largely adopted by national statistical offices because of the difficulty of assessing non-market output quantities and prices. Its use is now regulated by the European System of Accounts (Eurostat, 2010).

The first problem with this approach is that, as shown in the previous section, prices, marginal valuation and marginal costs usually do not coincide in the non-market sector. It would be preferred to evaluate every product at its user's marginal valuation (figure 2.2.1, panel B), nevertheless, information on marginal valuation is difficult to find, and statisticians have ended up employing the output=input method, acknowledging the high and likely risk of underestimating the total non-market production.

However, for the scope of this discussion, the biggest problem with setting the total value of output equal to the value of inputs is that the real volume growth of output is set equal to the real volume growth of inputs too. It can be easily derived that total factor productivity (TFP) will always be equal to one, meaning no positive or negative evolutions (Eurostat, 2016). This computation method reflects the view expressed by Kaufman (1976), which considers the public sector as being composed by static or unchanging organizations that cannot understand from the past and improve their productive processes. Nevertheless, Dunleavy (2017) recognizes that public bodies (as other market organizations) have been going through a great deal of structural changes, such as increased capital intensity and IT revolution, which makes it difficult to continue believing in this standstill theory.

First developments towards output-based methods took place in developed countries, with the Blue Book (HM Government, 1998) for the UK and Eurostat (2001) highlighting the importance of implementing direct quantity measures for non-market activities and giving the first guidelines[6]. Nevertheless, the discussion was by no means new: Hill (1977) had already stated the desirability of these methods for health and education services. In fact, these were the first sectors for

[6] Some forms of direct output measures had already been in place in the United Kingdom between 1950 and 1960. They were then heavily criticized for the arbitrary nature of the weighting system employed for their aggregation, and eventually abandoned (Levitt and Joyce, 1987).

which output-based methods were used, not only given their relative importance in the total Government expenditure but also for being individually provided services. Nowadays, almost all OECD countries have put in place direct measures of output for these services, although a standardized structure still does not exist. Schreyer (2012) offers an updated methodological framework to understand quantity and quality changes for these sectors, tackling the problem of quality adjustments, to which I shall come back on the following pages.

2.3.2 How to explain outputs?

When analysing companies that operate in the market sector, it is not difficult to identify their outputs: one just needs to take into consideration all the transactions among economic agents to understand which goods and services are exchanged in return for money. The lack of a similar mechanism makes the task more difficult in the Government sector, but defining a way to measure output for public organizations is the only way to implement productivity assessments. The aim of this section is to trace and summarize simple guidelines that should be undertaken to understand the ultimate question: 'what are public institutions producing?'. The outlined path will be useful in the third part of this paper, when discussing and applying this methodology for the two case studies of Dutch institutions.

As a starting point, it is convenient to refer to the Atkinson Review (2005), which outlines three pillars that should sustain every framework aiming to introduce direct measures of government outputs in the System of National Accounts:

- the design of direct output measures needs to be considered carefully and kept continuously monitored and updated, to adapt to changes in the production process and in the internal structure of the agency (e.g. services provided physically and measured by the number of clients entering a facility could switch to online, implying a change in the key performance indicator (KPI) considered). This implies that the process should not be understood as 'one-shot' implementation, requiring continuous investments of resources.
- the implementation and control should be independently carried out by third organizations.
- the output measures should reflect, as much as possible, the procedures adopted in national accounts for market output (private activities). Absence of compliance with this requirement may lead to changes in the national output only because of reallocation of activities (e.g. the privatization of a private school might lead to an increase in the national output, ceteris paribus)."

Dunleavy (2017), describing the process to identify output, introduces the concept of core output, highlighting that we should not consider a plethora of indices for each public institution, rather focusing on maximum one/two outputs for small agencies and ten/fifteen for large agencies with different tasks and objectives.

How to define core output and activities? The Atkinson Review (2005) provides more concrete and applicable guidelines, which suggest looking at households and firms as the final consumers of government services. Analysing their usage of public products and their interactions with government agencies, we should seek to identify the final services provided and, consequently, the data mirroring and recording these transactions. Therefore, it is clear that there is not a unique and universally applicable principle to pinpoint outputs, but a careful and close analysis of the activities carried out by each public institution is needed.

Although practical implementations of these guidelines are not easy to be put in place, measuring changes in the volume of output is not the only difficult task required to calculate productivity. Statisticians agree that dealing with quality changes is one of the most challenging aspects when measuring public institutions' outputs (Lau, Lonti and Schultz, 2017).

At this stage, it is clear that the main point to understand productivity is related to changes in the volume of output and inputs, disregarding changes in prices. Nevertheless, as fairly explained in the Handbook on Prices and Volume Measures in National Accounts (Eurostat, 2016), volume indices should capture changes in the:
- quantity of products
- aggregate consumption's composition (shifts between products)
- characteristics of the products

If we want to reach a reliable measure of output, we need to understand that prices can change from one year to the next as consequence of two different mechanisms: (i) variation in the characteristics of the product, and (ii) pure price changes. While the latter should be captured by the price index, the former effect needs to be registered by the volume index. To better grasp the importance of adjusting for quality, it is useful to consider a simple example: if labour productivity in education is calculated by dividing the number of pupils by the numbers of teachers, a reduction in the maximum number of students per class would be recorded as a reduction in productivity. However, it may be argued that a lower pupils-teacher ratio enhances a 'more personalized' educational system, increasing student preparation. Therefore, if developments in output standards cannot be registered as a decrease in productivity, quality needs to be accounted for. As acknowledged by international accounting rules[7] and national statisticians, consumers do not only benefit from the quantity of products they buy but also from their nature, valuing quality improvements[8] (Productivity Commission, 2017). Moreover, although the problem of quality evaluations affects both market (e.g. IT products[9]) and non-market sectors, the severity of the concern is higher in the latter, since services are relatively more offered than physical goods[10] and prices, which could somewhat represent quality levels, are absent (Schreyer, 2010).

In a market context, the issue can (partially) be solved using information about transactions; Eurostat (2016, pages 20-21) explains that consumers' preferences are revealed by their purchasing behaviours, and continues:

7 Eurostat (2016): "Changes in quality over time need to be recorded as changes in volume and not as changes in price."

8 Same reasoning can be applied to the health sector: more surgeries (with the same level of inputs) would be registered as an increase in productivity, but also improvements in the survival rate or fewer recovery days should be recorded as productivity improvements.

9 See Triplett (2006).

10 Because of their inner characteristics, quality in services is more difficult to be measure than in physical goods (Dunleavy, 2017).

"A difference in price that exists between two products at the same time can be interpreted as the value that consumers attach to the quality difference between the two products. This implies that a higher price is associated with a higher quality. If shifts in the quantity of consumption occur between the different products this should be seen as a volume change, implying that the quality difference between the two products is exactly equal to the price difference."

This brings us to the first strategy that can be employed to deal with quality changes: product differentiation. Researchers agree that part of the quality adjustments can be accounted for by differentiating as many separate products as possible; that is, products that differ in terms of quality are treated as separate items (Schreyer, 2010). It follows that if consumers shift towards a higher-quality product, this will be recorded as a compositional change in the aggregate. In the non-market context, prices are not available and the Atkinson Review (2005, page 91) suggests substituting them with costs: *"Suppose that there are two treatments. One is of a higher quality and is more expensive. If the output indicator combines the two volume measures with weights according to their cost, then a shift towards increased use of the higher quality treatment will be properly recorded. There will be an increase in expenditure, and a corresponding increase in output. Thus, the move to a more detailed treatment classification has been a step towards taking account of this kind of quality change."*

Nevertheless, one important caveat of this approach is its lack of adaptability to the development of more efficient technologies, which are cheaper than the incumbent ones. In this case, using cost weights, the improvement will be registered as a reduction in output and, consequently, no changes in productivity, while, in reality, a positive shift happened (see Schreyer's (2012) page 268).

Schreyer (2012) discusses this last issue and highlights the importance of using a correct level of product differentiation. If consumers consider the two treatments (old and new technology) as perfect substitutes, no adjustments (cost weighting) are required, and the total volume index would simply be calculated adding up the number of products produced with the old and new technology. This brings the analysis of what consumers value to the discussion, that is, every analysis about quality adjustments should, to some extent, consider outcomes, even if only to establish product differentiation criteria. Although this strategy works in theory, one strong initial assumption makes it difficult to be applied in real terms: products quantities can be summed up without adjustments only if consumers are perfectly indifferent between them. In reality, consumers rarely do not have preferences when choosing among different qualities: if we reconsider the healthcare example, patients will probably prefer the new and perhaps less invasive technologies rather than the older ones.

Consequently, when implicit quality adjustments (product differentiation) are not enough to account for all quality differences, explicit adjustments are necessary (Schreyer, 2010). One of these adjustments can be applied re-scaling products' quantity indices and expressing them in relative terms. In practice, either the volume of the innovative product can be increased, or the traditional one can be decreased to account for purchasers' preferences. Clearly, the absence of data about consumers' preferences makes this approach difficult to be put in place.

A second explicit adjustment involves the use of hedonic price indices, which are calculated employing hedonic functions. The idea is to regress the price of different items against the quantity of their attributes (Triplett, 2006). In this way, the marginal impact of every feature is estimated, and products' quality can be expressed as a function of the volume of their characteristics. This method is successfully applied for some market sectors (e.g. real estates and IT products), but is often overlooked for non-market contexts. Schreyer (2012) demonstrates its applicability for public goods, though highlighting that its implementation is undermined by many empirical problems (e.g. data availability, choice of characteristics to be regressed). He concludes that more research and exploratory applications are necessary before clearly assessing its usefulness in the non-market context.

A third approach calls for calculating "the effect of the (public) service on the marginal outcome of the consumer, particularly with relation to its degree of success" (Productivity Commission, 2017). Atkinson (2005) argues that this approach may consist in a direct way to measure outputs in the non-market sector, which would not require any other explicit quality adjustments. In practice, changes in outcome levels would be tracked and related to changes in the output level. First applications of this strategy were carried out for the education sector, with Jorgenson and Fraumeni (1989) analysing the marginal addition of education on human capital, while Cutler et al. (1998) analysed the marginal impact of heart attack treatments. Nonetheless, it should be made clear that, when considering outcome, its value is not only influenced by the output we want to analyse, but also by other factors[11]. For example, the national health level (outcome) is influenced by the health system quality (output) but also by other variables such as alimentation, sports, lifestyle, etc. Indeed, there are some empirical issues that make the application of 'marginal calculations' difficult to realize, such as: (i) it is difficult to control for all the factors influencing outcome other than the output under analysis, (ii) results are biased by the individual capacity of the consumer to make use of the products[12] (Schreyer, 2012). Furthermore, in a recent work, Cutler et al. (2020) undertake a related empirical analysis for the health sector in the U.S., proposing a satellite account. Their innovative feature is to consider 'medical conditions' as the criteria to individuate different "industries" (e.g. heart diseases, brain cancer, etc.). Within this framework, changes in spending for each 'industry' are correlated with changes in the overall health status of people affected by that specific disease. In other words, to measure productivity, they only rely on an explicit connection between input (medical spending) and outcome (health situation).

The bottom line is that there is not a unique and preferred method for quality adjustments. Each case should be considered apart, given its peculiarities and data

[11] For this reason, we should be keen on understanding output's marginal impact on outcome.
[12] As an example, the very same teaching activity (lesson) given by a teacher can have different impacts on student with diverse backgrounds, with some benefitting more than others.

availability. In general, some references to outcomes need to be made, but it is still uncertain until which point. Researchers and statisticians agree that more work on this topic needs to be carried out before coming up with solid and internationally recognized methods. The Atkinson Review (2005) states that priority needs to be given to research on quality adjustments; on the other hand, until there is strong consensus about their reliability, they should not be included in the National Account calculations.

2.4 Valuing productivity when information on output quantities is available

Once information on real outputs and inputs is available, it is possible to calculate productivity, which has been defined as the ratio between outputs and inputs (see equation (1)).

To do so, some methodological considerations and assumptions about the production function should be given. According to Hulten (2010), an atheoretical approach to TFP calculation cannot give solid criteria for any type of growth account. To this extent, Solow (1956) proposed his well-renown explicit production model based on a Cobb-Douglas equation, where output (Y_t) is the product of a function of inputs ($F(\cdot)$) and a factor-neutral shifter[13] (A_t), as shown in equation (3).

$$Y_t = A_t F(K_t^{\alpha} L_t^{\beta} M_t^{\gamma}) \tag{3}$$

A_t (or TFP) is usually referred to as the 'measure of our ignorance', since it is determined as the residual and it captures the variation in output that cannot be explained by the observable inputs contained in $F(\cdot)$, which are capital (K_t), labour (L_t), and intermediate material (M_t), α, β and γ being the respective elasticities of substitution. That is, movements on the production function are gauged by changes in the quantity (and composition) of inputs, while the residual (A_t) represents production curve shifts.

For the Solow model to be reliable, several assumptions need to be made: (i) there is a stable relationship between inputs and outputs at the level of aggregation considered, (ii) neoclassical smoothness and curvature properties, (iii) inputs are paid their marginal value and are perfectly mobile, (iv) there are constant returns to scale, (v) technical changes have a Hicks'-neutral shape (they do not affect the relative quantity of labour and capital in the production function).

Nevertheless, since information about the actual shape of the production function and the degree of compliance with the previous assumptions are not available, there is the need for more flexibility. One general function that allows some

13 Hulten (2001) states that TFP "*measures the shift in the production function*" (p.40).

relaxations[14] is the so-called translog production function, which is represented by equation (4).

$$lnY = \beta_0 + \sum_{i=1}^{n} \beta_i lnx_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{ij} lnx_i lnx_j \quad (4)$$

The aim of this formulation is to allow for non-constant elasticities of substitution for each input x_i through its interaction with each other possible input x_j in the production function. For example, assuming that x_i is capital, its impact on output is determined by β_i (the fixed elasticity considered in the Cobb-Douglas structure) plus the interaction of capital with every other input (the second term on the right side of equation (4)).

According to Hulten (2010) based on Diewert (1976), Törnqvist-Theil indices are exact when the underlying production function has a Translog shape. Therefore, following Diewert's (2017) approach, given that p_i^t and x_i^t represent respectively, inputs price and quantity vectors for $t = 1, \dots, t$, the period t Törnqvist Price Index[15] is, as shown in equation (5)[16].

$$P_T^t = exp\left[\sum_{i=1}^{I} (\frac{s_i^1 + s_i^t}{2}) \ln(\frac{p_i^t}{p_i^1})\right] \quad (5)$$

where $s_i^t = x_i^t p_i^t / x^t p^t$. Consequently, the Törnqvist Quantity Index can be calculated dividing each period's total input costs by the respective price index. In other words, applying equation (6), we are deflating aggregate cost figures.

$$Q_T^t = {\sum_{i=1}^{n} p_i^t \, x_i^t} \Big/ {P_T^t} \quad (6)$$

Without any information about real outputs, changes of input volume indices over time (of the type in equation (6)) are used to compute Government output real growth, according to the output=inputs convention (Diewert E. W., 2017; Atkinson, 2005). However, since now we are considering a situation in which information about output quantities (q^t for $t = 1, \dots, t$) is available, we can proceed calculating the normalized output (Q^t for $t = 1, \dots, t$).

$$Q^t = \sum_{i=1}^{n} p_i^1 \, x_i^1 \left[\frac{q^t}{q^1}\right] \quad (7)$$

14 Mainly elasticities of substitution and elasticity of scale are allowed to change with output and input proportions (Heatfield and Wibe, 1987.

15 Alternatively, other price indices can be calculated and used in the TFP computation. Following Diewert (2017), the period t Paasche Price Index is $P_P^t = {\sum_{i=1}^{n} p_i^t \, x_i^t} \Big/ {\sum_{i=1}^{n} p_i^1 \, x_i^t}$.

16 Note that with this notation (fixed base index), the first year (t=1) is considered as the base year. Alternatively, Diewert (2017) proposes the use of chained indices. This implies that for the first two year the chained indices will be equal to the fixed ones, while from the third period they will be equal to the period t-1 chained index level times the Paasche rate of change of input prices from period t-1 to t.

This is necessary to apply the input indices developed before, since output in period 1 needs to be equal to the total inputs cost in the same period[17]. Consequently, productivity can be calculated as the ratio between the two quantity measures (inputs and outputs):

$$TFP^t = \frac{Q^t}{Q_T^t} \tag{8}$$

2.5 Issues on the aggregation of outputs

When a public body produces only one output (good or service), total factor productivity can be calculated simply dividing the number representing the volume of output by the one representing the volume of the inputs used in the production process. However, when the agency delivers more than one good or service, we need to find a proper way to combine these volume measures into an aggregated quantity index, using the right weights.

One of the three Atkinson's pillars (Atkinson, 2005) suggests looking at the market sector: there, prices (as they are representative of users' marginal valuation) are used as weights to aggregate different products. Schreyer (2012, page 7) affirms that "*when weights are needed to aggregate across products, there is no need to invoke either a consumer or a producer perspective — the value of market transactions is all that is needed, and it combines the two sides of the market*". Therefore, goods and services for which consumers are marginally willing to pay higher prices receive a higher weight in the national income than products for which consumers would pay a lower amount of money. Nevertheless, although this explanation gives some fruitful insights on how prices should be interpreted, it cannot be used to find weights for aggregating different government outputs, since there are no transactions from which information on quantity and prices can be obtained.

Douglas (2006), in a work commissioned by the New Zealand Treasury, argues that in the absence of consumers' marginal valuation, prices for the same activities and offered by the private sector should be employed[18]. Nonetheless, the situations where information from the private market can be used for non-market output are limited (e.g. there is not a private substitute for tax administration offices), so that the initial question about how to aggregate output remains open for further research. Douglas (2006) recognizes that, in practice, both the methods explained above (marginal valuation and market prices) are difficult to be applied, so that *cost weights* are often used in aggregating output, and states:
"This means that each different type of output is weighted by the cost of providing that output before the outputs are added together. By doing this, we are using the relative per-unit costs as a proxy for the relative per-unit values to the consumer. In

17 In particular, this notation is important to calculate the cost-based output price index: $P^t = \frac{\sum_{i=1}^{n} p_i^t x_i^t}{Q^t}$

18 For example, to evaluate the output of road infrastructures, where we would need to sum up in the same indicator passengers and freights, the alternative costs of using rail transportations might be useful weights.

a competitive market, where all output is allocated until marginal cost is equal to marginal value, this may be valid. However, for a public service that might be under or over allocating services, this is not ideal and should be avoided if actual value weights are available."

The above discussion can be summarized as follows:

1. outputs should be aggregated according to their marginal values (equal to prices, for market transactions);
2. in the absence of meaningful price and marginal valuations information, if similar products are offered by market institutions, their price information should be used;
3. if none of these two options is available, cost weights are applied[19].

Nevertheless, we should be extremely clear on the implicit assumptions we make when using cost weights as a way to aggregate outputs, since we suppose producer's cost information to be representative of the value consumers attach to each product. This would be the case in private organizations operating under a competitive scenario, where their management teams steer the activity to meet customers' desires, shifting the cost structure consequently. For example, in a two-product world, if consumers grow to prefer Product A over Product B, firms will shift resources from the production of the latter to the former. As a consequence, the share of costs relative to Product A will increase, while the other will shrink. If the same assumption applies to Government institutions, so that their activities are steered to meet citizens' preferences, cost weighting can be applied effectively and no further considerations about policy objectives are needed.

At this point, the key enquiry is whether public institutions can be considered the same as private ones, operating under market rules. According to Rosenberg Hansen and Ferlie (2016), there is a well-established literature that considers the nature of management practices in public organizations significantly different from the ones of private companies. There may be two potential reasons: the more unclear and complicated goals of the public sector and the broader and the politically driven decision process (Rainey, 2009). On the other hand, public organizations in various countries have been going through reforms, especially since the 1980s (New Public Management[20]), which according to Rainey and Chun (2005) have made researchers and experts agree on a more blurred distinction between private and public sectors, highlighting more similarities than differences. The extent to which we accept this last view determines the level of reliability of cost-based weighting systems.

To make matters more complicated, cost weights methods may also be difficult to be practically implemented. In the case of a single public agency producing more than one good/service but without a system of costs apportionment, it would be

19 For example, education is usually measured by the number of equivalent pupils, aggregated using total national spending figure each type of school (Simpson, 2009).

20 For further information see Dunleavy and Hood (1994).

impossible to understand which resources have been used to produce one output or the other .

2.6 Recent research in The Netherlands

Probably the most comprehensive research on the productivity of the Dutch government is done by the Institute for Public Sector Efficiency Studies (IPSE). They published research on, among other, the productivity in the field of security and justice (Blank and Van Heezik, 2019, 2020, 2021), the local administrative public sector (Blank, 2018) and on individual government agencies like the cadastre (Niaounakis and Van Heezik, 2020) and the social benefits agencies (Niaounakis et al., 2020).

Moolenaar (2019) of the scientific research and documentation centre (WODC) investigated the possibilities to calculate the development of the (labour) productivity of all organizations within the Ministry of Justice and Security for the period 2013 - 2017.

In 2017, the 'Centraal Planbureau' (CPB) published a note on the productivity growth of the Dutch Government as a whole (CPB, 2017). This research was commissioned by then Ministry of Internal Affairs and Kingdom Relations. The year after, this note was followed by a publication (CPB, 2018) in which two different method to measure the productivity of Government were compared. More recently Mellens et al. (2020) looked into the factors that affect the productivity like education, training and the use of ICT.

3. Data Analysis and Case Studies

Given the previous methodological background, the remaining part of this paper considers two specific public functions in The Netherlands: (i) the body of public libraries and (ii) the fire department. Firstly, a discussion about inputs, their data sources and deflation is proposed; secondly, for each of the two institutions, their productive processes are analysed to individuate output measures and estimate productivity. Finally, some recommendations for improvements in the data collection process are put forward.

3.1 Inputs

This analysis' starting point for considerations about inputs attribution is the COFOG classification, which is the international standard for the codification of Government activities based on the "*purposes for which the funds are used*" (OECD, 2017). In other words, it represents an industrial classification system that groups public spending into homogenous activities. This implies that using COFOG in this analysis is particularly useful, as it makes easier to attach the right set of outputs to each set of inputs. That is, Government expenditures are classified in 12 categories (one-digit level), which in turn are divided in sub-sections (two-digit level) and for each of these sub-categories, costs are classified by their nature. In fact, it is worth noting that Government expenditure includes also distributive transactions (e.g. social benefits and transfers in kind), whose magnitude does not affect the productive process itself. Table 3.1.1 shows the transactions that need to be aggregated into a (nominal) measure of input; data for The Netherlands are made available on StatLine at CBS.

Table 3.1.1. The calculation of total nominal inputs

Code		Transaction	Source
P.2		Intermediate consumption	
D.1	+	Compensation of employees	
D.29	+	Taxes on production	(CBS, 2020a)
D.39	-	Subsidies on production	
P.51C	+	Depreciation of fixed capital	
		Total (nominal) inputs	

The next step is to move from nominal cost figures to their volume counterparts. Two strategies can be employed to obtain inputs quantity indices: (i) the use of price indices to deflate nominal values, (ii) the direct measure of changes in the volume of inputs. In practice, while the former can be applied for all types of transaction, the latter can be used only to calculate labour inputs, as information about the number of fte or the number of hours worked is usually readily available.

As stated in the Atkinson Review *2005, page 51), a general requirement for deflators to be reliable is that: "*they should be sufficiently disaggregated to take account of changes in the mix of inputs and should reflect full and actual costs*". Therefore, the evolution of different price categories can be used to deflate different transactions, in The Netherlands:

- The price index of the material consumption of government (IMOC) may be considered to represent the evolution of material consumption prices in Government expenses and can be employed to deflate intermediate consumption figures.
- The price of labour for the Government sector can be used to derive an indirect measure of the volume of labour inputs, since it reflects the evolution in the average compensation of public employees. The index is calculated by weighing the development of the wage costs per hour worked in different labour categories (classification accounts for gender, age group, education level, industry). Note that this index is calculated at the industry level (in this case, public administration and government services), meaning that institution-specific changes in the labour structure will not be accounted for. For this reason, direct measures of employees' volumes are usually preferred, as it will be shown in the case of libraries. However, for the fire department, the lack of meaningful volume measures imposes the usage of indirect deflation.
- The price index of gross government investments (IBOI) tracks price movements for gross Government investments and, therefore, it can be used to deflate fixed capital consumption.
- the consumer price index (CPI) is employed to deflate taxes and subsidies.

Note that these measures could be furtherly refined and improved to be applicable in the analysis of specific public organizations; for example, the index of the price of labour might be weighted by education and skill levels that better reflect the composition of specific workforces. Table 3.1.2 summarizes these indices and their sources.

Table 3.1.2. Price indices for Government expenditure in The Netherlands

Transaction	Index	Source
Intermediate consumption	IMOC	(CBS, 2019a)
Compensation of employees	PvA	(CBS, 2020b)
Taxes on production	CPI	(CBS, 2020c)
Subsidies on production	CPI	
Depreciation of fixed capital	IBOI	(CBS, 2020d)

How are these numbers used in practice? Firstly, input value figures are indexed to a reference year. Secondly, annual price changes are chained together to obtain an index (the reference year must be the same as the one used for nominal values) and used to deflate singular transaction categories. Finally, the former elements are divided by the latter and summed up according to their share in the total expenses of the previous year. Further issues about capital and labour costs computation (the opportunity cost of capital and quality-adjusted labour volume) are tackled in the appendix.

3.2 Public Libraries

The first data analysis takes productivity changes in the system of public libraries into consideration. In the following paragraphs, a careful assessment of libraries' productive process is proposed and employed to derive an output index, which, in turn, fuels a sensitivity analysis aimed to better understand TFP variation overtime. Finally, the potential impact of volunteers' underestimation is calculated.

3.2.1 Framing the Context

To understand and measure the processes taking place within the system of public libraries, the analysis should start from its policy foundations. Essentially, interpreting policymakers' decisions and implementations gives information about the scope and the objectives the organization is pursuing in the society.

In the last decades, public libraries in The Netherlands have been going through a period of great reforms and reorganizations. The first step towards the current system of public libraries was set in place in 1998, with the Dutch Council for Culture recommending a more cohesive and uniform national structure (Raad voor Cultuur, 1998). According to Huysmans and Hillebrink (2008), this triggered a long process of mergers and consolidations that led to an overall reduction in the number of independent library organizations and facilities. In 2005, the Netherlands Public Library Association (VOB) published the 'Richtlijn voor basisbibliotheken' (Guideline for basic libraries), which reported the core functions that the library sector set out for itself (Vereniging van Openbare Bibliotheken, 2005). The current legal framework came into force at the beginning of 2015 with the Public Library Facilities System Act (WSOB). Building on the previous structure, (i) it better defines libraries' social function, (ii) it fosters cohesion and aggregation, and (iii) it organizes the digital domain. The first point is summarized in five main objectives libraries should address:

1. Making knowledge and information available (informing)
2. Providing opportunities for development and education (learning)
3. Promoting reading and introduction to literature (reading)
4. Organizing meeting and debate
5. Introducing art and culture

Furthermore, to better approach the second and third points (ii, iii), the act has connected the system of public libraries with the Koninklijke Bibliotheek (Royal Library), giving it a coordinating role and responsibility for the national digital library, which is now a pivotal part of the network. The WSOB has also led to the creation of a special research department (operating within the legal framework of the Koninklijke Bibiliotheek) for the collection and analysis of the data recorded by each single library: the Bibliotheekinzicht.

3.2.2 Output Data

Since 2015, the Bibliotheekinzicht has been delivering annual reports aimed at describing the national network of public libraries, giving precious information about its productive process and the services offered. This data, coupled with the ones made available by Statistics Netherlands on StatLine, give a broad perspective on the developments taking place within public libraries. Table 3.2.2.1 presents all the variables used in the following process of outputs calculation and aggregation.

Table 3.2.2.1. Libraries' output indicators

	2015	2018	Average (2015-18)
Number of public libraries, $N.LIB._i$	156	146	151
Physical collection (x 1000), $\boldsymbol{COLL_i}$	25 356	24 252	24 917
Physical loans (x 1000), $\boldsymbol{PL_i}$	78 069	66 537	72 391
Loans, downloads and use of digital material (e-book, audiobook and courses) (x 1000), $\boldsymbol{DL_i}$	4 035	6 588	5 514
Digital collection (x 1000), $\boldsymbol{LIC_i}$	10.68	23.90	16.81
Annual visits (.000), V_i	55 632	62 649	59 528
Number of informing activities[21]	4 028	14 137	8 913
Number of learning activities[22]	24 991	87 041	49 909
Number of reading activities[23]	36 236	65 662	51 219
Number of meeting and debates[24]	4 946	8 732	6 328
Number of art and culture activities[25]	8 615	26 451	14 445

Sources: CBS (2019b, 2019c), Bibliotheekinzicht (2020)

3.2.3 Process Analysis and Productivity Calculation

The previous brief exploration gives two important pieces of information that are crucial to structure considerations about outputs. Firstly, there has been a shift in the role of public libraries: from simply ‘making knowledge available’ to a more holistic function. In fact, the core passive activities, such as lending books and providing facilities for self-studying, have been integrated with a range of services that aim at actively involving citizens. Secondly, digitalization has radically impacted and changed the way people make use of media and information material (Huysmans and Hillebrink, 2008). Libraries are trying to adapt to these developments, as the implementation of a centralized national digital library is at the centre of their strategy. Consequently, information about collections, loans and users alone cannot be fully representative of the productive process of public libraries. Therefore, the overall output measure should, to some extent, encompass the digital sphere and the one relative to events/ meetings.

A first attempt to evaluate libraries’ outputs has been carried out by the Kwink Groep (2019). However, since the final objective of their report was the

21 Activities aimed at information and knowledge sharing, such as information meetings on topics such as work and income, health, legal, society and science (Kwink Groep, 2019).

22 Activities aimed at the development of information literacy and media literacy for primary and secondary education, activities for adults in the field of language and digital skills (Kwink Groep, 2019).

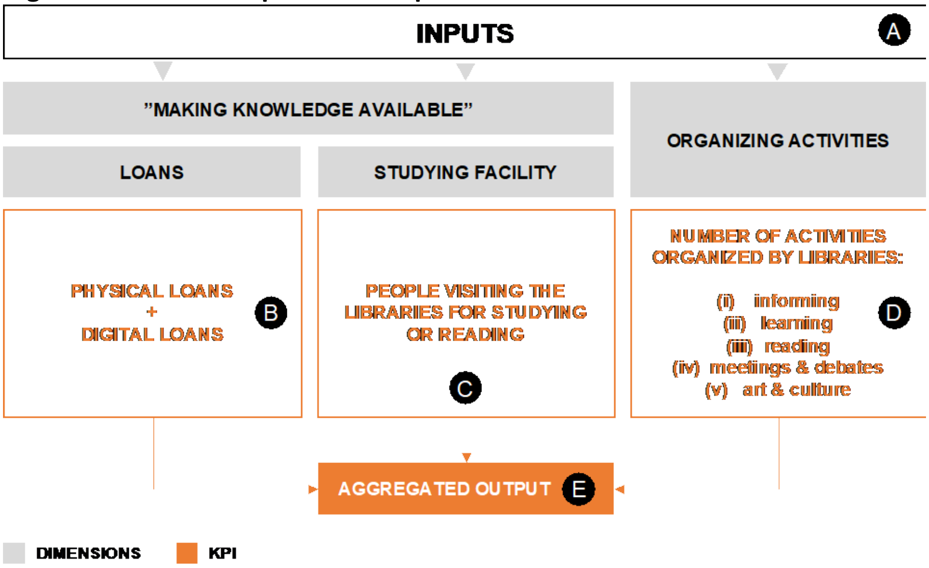
23 Activities related to reading promotion for primary and secondary education, pre and early childhood education, extra-curricular reading activities and reading clubs (Kwink Groep, 2019).

24 Activities aimed at meeting (e.g. walk-in coffee, theme café, lunch, quiz, market) and/ or discussions about social themes (Kwink Groep, 2019).

25 Activities such as lectures, exhibitions, courses, workshops, theatre performances, film screenings, concerts, open stage (e.g. poetry, singing), art routes and city walks (Kwink Groep, 2019).

assessment of the new reform implemented in 2015, they did not analyse productivity and focused prevalently on the “activities side” of production. This approach discards the important role of the fundamental services offered by libraries (borrowing books and making studying facilities available). Therefore, as a comprehensive approach to aggregate outputs needs to be undertaken, Figure 3 shows the three main dimensions of libraries’ production and the (potential) related KPIs. The challenges/issues that need to be tackled before obtaining a final index (the letters in the picture) are discussed on the following pages.

Figure 3.2.3.1. The productive process of libraries



Issue A – Inputs Data

Previously, it has been stated that the starting point for considerations about the input side is the COFOG dataset. Nonetheless, this classification is compiled up to the second-digit level, while data about libraries alone could only be found in a hypothetical third-level segmentation. Therefore, the following productivity measures rely on costs figures made available on StatLine (CBS, 2019b) thanks to a collaboration with the Bibliotheekinzicht. This dataset follows a different classification method[26], so that its costs items need to be reconciliated with the aforementioned three-dimensional structure composed by labour, capital and intermediate consumption. Panel A of table 3.2.3.2 shows the different costs items for the years 2015-2018, expressed in current values, Panel B reports direct measures of volume for labour expenditure.

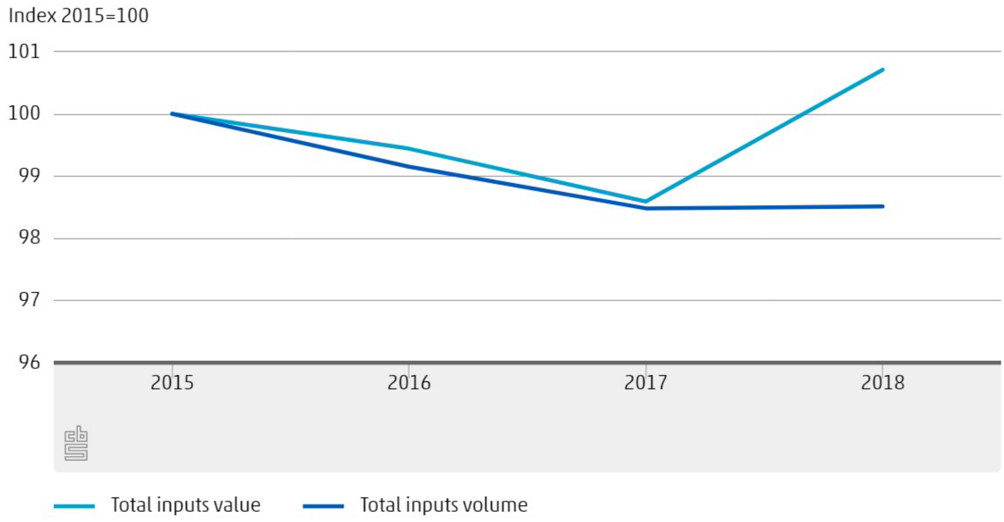
[26] Charges are divided in these categories: Housing costs, Salaried Personnel, Personnel not employed, Administration and automation, Media Costs and Other Costs.

Table 3.2.3.2. Inputs current value, deflators and direct measures of inputs volume

	2015	2018	Source
PANEL A			
Intermediate consumption (mln euro)	272.50	255.50	
Compensation of employees (mln euro)	227.70	244.90	(CBS, 2019b)
Compensation of volunteers (mln euro)	17.30	18.90	
Capital depreciation (mln euro)	15	17	CBS[27]
PANEL B			
Total employees (fte)	4 143	4 213	(CBS, 2019b)
Total volunteers (headcount)	10 828	19 776	

Except for employees' and volunteers' compensation, all other cost items are deflated using the respective price indices reported in table 3.1.2. The former ones are directly measured in volume terms: the number of FTE for salaried personnel and a simple headcount for volunteers. [28] However, since not only the volume but also the average worker's skill level might have changed, this methodology does not account for potential quality modifications in the labour force. Unfortunately, detailed data about employees' wages are not available, so that quality-adjusted weighting systems are not implementable. Deflated volume measures for intermediate consumption, labour (paid and non-paid staff) and capital are aggregated according to the geometric mean, where the exponents (i.e. weights) are the simple average of their share in the total costs in current and past year. Figure 3.2.3.3 depicts the development of nominal inputs and the final index of deflated inputs (dark blue line).

Figure 3.2.3.3. Nominal Inputs and real inputs

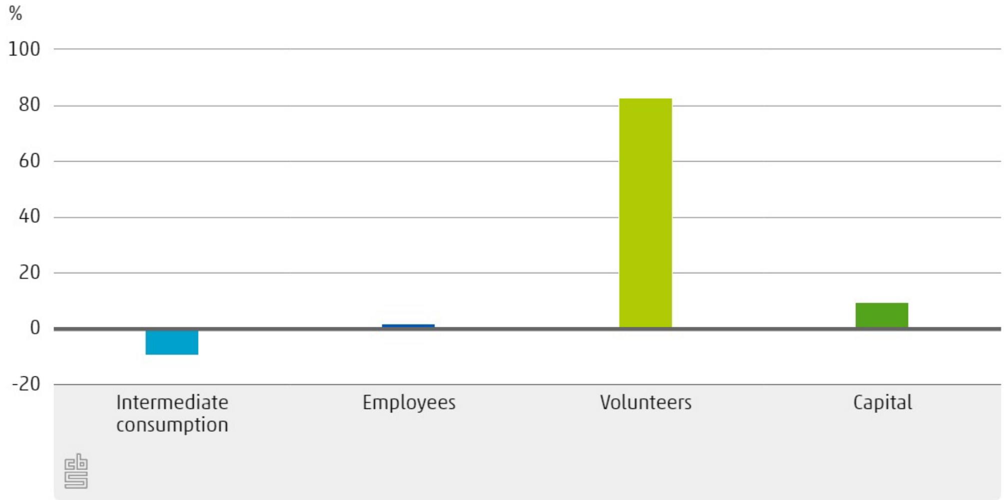


Although the figure of inputs volume seems rather stable overtime, figure 3.2.3.4 shows the changes between 2015 and 2018 in its four main components: intermediate consumption, employees' salaries, non-employed workers'

27 These data are not available in StatLine.
28 As FTEs information is not available for volunteers, using a simple headcount implies the assumption that hours worked per volunteer remain constant over time

compensation and capital expenditures. The graph highlights that a decrease in intermediate consumption was compensated by an increase in the volume of salaried staff, volunteers and capital (the latter two having a lower impact, according to their average shares in the total costs).

Figure 3.2.3.4. The components of inputs growth and their share, 2015-18



Note that the aggregated inputs index is calculated according to the different components' cost shares. Nevertheless, this approach is correct only under the assumption that each factor is paid at its marginal productivity, which might not be true for volunteers. Some considerations on their impact on the final TFP are proposed at the end of this section.

Issue B – Aggregating loans data

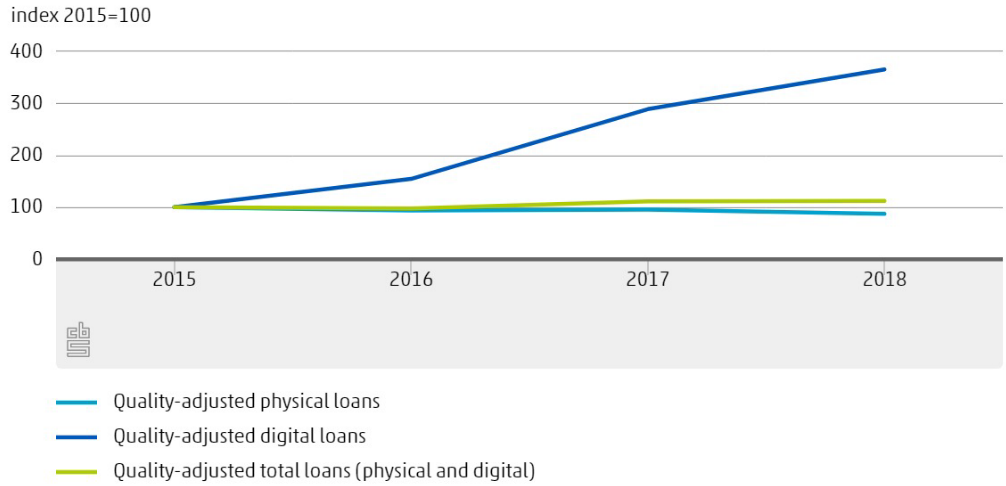
The first dimension of libraries' production is loans, both physical and digital. Data about these two aspects needs to be adjusted by quality changes and then, aggregated. Firstly, the volumes of "borrowing activities" need to be adjusted by their quality: a library that gives the possibility to choose a book among a total collection of one thousand titles is providing a worse service that one offering more than a million references. This follows the results proposed by Brynjolfsson, Hu and Smith (2003), who argue that increased product variety (book titles) has a large impact on the consumer surplus gains, even bigger than the one generated by improved efficiency (more competition and lower prices).

Consequently, the volume of loans is adjusted by (i) the average collection per library, for the physical ones; and (ii) the number of active licenses, for those borrowed via the Internet (dashed lines in figure 3.2.3.5. In practice, the two indices (loans and collection) are multiplied, so that if the former increases while the latter declines, the final indicator will take both the aspects into account (see equation (9)).

Nevertheless, it is worth emphasizing that these adjustments implicitly infer the existence of a specific production function with equal weights to additional loans and improved library collection. That is, the marginal rate of substitution between these two dimensions needs to be better determined to answer the questions: how does the additional availability of books impact on consumers' service evaluation? How does this effect compare with additional volume (loans)? Then, the two quality-adjusted measures need to be aggregated. Since no price information is available, this analysis considers one book borrowed in a library to have the same value (to the final user) as the same book downloaded/ read via digital devices. That is, as shown in equation (9), the two adjusted indices are aggregated according to their share on the total number of annual loans (s_i^P for physical loans, and $s_i^D$for digital ones), where PL_i is the annual volume of loans, $COLL_i$ the total physical collection, $N.LIB._i$ the number of active libraries, DL_i represents the annual digital loans, and LIC_i is the annual measure of active digital licences (titles available online).

$$L_i = s_i^P \left(PL_i * \frac{COLL_i}{N.LIB._i} \right) + s_i^D (DL_i * LIC_i) \quad (9)$$

Figure 3.2.3.5. Quality-adjusted physical, digital and aggregated loans



Issue C – Estimating the number of people visiting libraries for studying
A second function responding to the final objective of making knowledge available is giving people the possibility to use libraries' facilities for activities such as self-study, reading or using computers and other technological devices. Although the Bibliotheekinzicht has been estimating the total number of visits since 2005 (V_i), unfortunately, there is no information about the actual motives why people frequent these facilities. In other words, based on the data currently available, there are no ways to derive a volume indicator for the second dimension mentioned in figure 3.2.3.1.

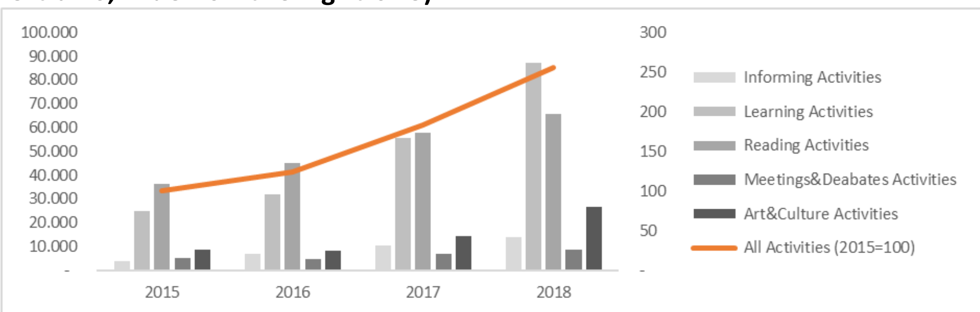
Nonetheless, note that we are interested in the evolution of these numbers over time and not in their absolute values, so that a solution may entail the use of the aggregate change rate to describe the evolution in the number of people frequenting libraries to study and read. It is worth emphasizing that this is a strong assumption, since it may be the case that visits for "events" increase (iii) while the ones for the mere use of facilities decrease (ii). Moreover, it could be argued that the number of loans may be used to estimate the number of visits related to this dimension; however, without information about the average number of books (loans) per visit, this estimation cannot be applied. Therefore, as it will be reported on the following pages, one way to improve the model reliability is to collect more detailed data for library visits.

Issue D – Aggregating different activities
The third function of libraries, which is emphasized in the WSOB and by the Kwink Groep's report (2019), pertains the organization of activities/events. Their annual volume is recorded by the Bibliotheekinzicht, which classifies them in five categories, according to the five pillars of libraries (see before). Consequently, since these numbers need to be summed up into a unique index, the following model gives each activity (a_i^n, where *i* is the year and *n* the specific group of activities) an equal weight (see (10)).

$$A_i = \sum_{n=1}^{N} a_i^n \tag{10}$$

This assumption is based on the principles expressed by the WSOB in 2015, which does not attach different importance to the five objectives mentioned before, so that each one should be pursued with the same intensity and effort. However, also in this case, the aggregation is quite arbitrary, as it may be the case that activities in one of the five categories are more valued by final users or more costly for the library system. Activities and events are usually organized without charging participants, so that price information cannot be registered. However, the second-best option to improve this weighting system entails recording the detailed costs (including labour) sustained in the organization of each activity. Figure 3.2.3.6 displays the development the absolute number of activities organised (grey bars) and the aggregated index (orange line).

Figure 3.2.3.6. Number of activities organised (absolute numbers on the left axis, index on the right one)



Issue E – Combining the three dimensions: TFP sensitivity analysis
The numbers representing total loans (L_i), the use of libraries' facilities (V_i), and activities (A_i), which reflect the three dimensions of the productive process depicted in figure 3.2.3.1, are then indexed to the same reference year (2015=100). They are reported in the first part of table 3.2.3.7. The last and most challenging issue is their aggregation into a single index (O_i), which needs to reflect the total production of libraries. This calculation is shown in Equation (11), where $w_L, w_V, and\ w_A$ represent the weights attached to the three dimensions mentioned above (note that their sum is 1).

$$O_i = w_L * L_i + w_V * V_i + w_A * A_i \qquad (11)$$

In fact, the challenge is to determine the right weight to attach to each indicator, since no information about users' valuation, private sector or cost structure is available. If a more detailed dataset about costs was implemented, one could use it to understand how libraries are investing their resources and, under the assumption that their management steers the activities to meet customers' requests, this figure would somehow reflect users' marginal valuation. Nevertheless, since this is only a speculation, the discussion proceeds relying on a sensitivity analysis, which allows to make considerations about the evolution of productivity, while varying the weights of the different dimensions. That is, for each year, different combinations of weights are employed to obtain output estimations, which are then divided by the index representing inputs to calculate productivity. Table 3.2.3.7 shows this calculation in three different scenarios, highlighting how determining weights has a substantial impact on the final productivity level (+24% in scenario 1 versus +59% in scenario 3).

Table 3.2.3.7. Libraries' output indices and TFP calculation

	2015	2016	2017	2018
Inputs	100	99	98	99
Loans **(L)**	100	97	111	112
Visits **(V)**	100	108	107	113
Activities **(A)**	100	124	184	256
Scenario 1 (w_L=80%; w_V=10%; w_A=10%)	100	102	118	124
Scenario 2 (w_L=60%; w_V=10%; w_A=30%)	100	107	131	146
Scenario 3 (w_L=40%; w_V=20%; w_A=40%)	100	111	137	159

Figure 3.2.3.8. 2018 TFP Sensitivity Analysis (2015=100)

Loans \ Activities	0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%
0%	114	119	124	129	135	140	146	152	159	166	172	180	187	195	203	212	221	230	240	250	260
5%	114	119	124	129	135	140	146	152	159	165	172	180	187	195	203	212	221	230	240	250	
10%	114	119	124	129	135	140	146	152	159	165	172	180	187	195	203	212	221	230	240		
15%	114	119	124	129	135	140	146	152	159	165	172	180	187	195	203	212	221	230			
20%	114	119	124	129	135	140	146	152	159	165	172	180	187	195	203	212	221				
25%	114	119	124	129	135	140	146	152	159	165	172	180	187	195	203	212					
30%	114	119	124	129	135	140	146	152	159	165	172	179	187	195	203						
35%	114	119	124	129	135	140	146	152	159	165	172	179	187	195							
40%	114	119	124	129	135	140	146	152	159	165	172	179	187								
45%	114	119	124	129	135	140	146	152	159	165	172	179									
50%	114	119	124	129	134	140	146	152	159	165	172										
55%	114	119	124	129	134	140	146	152	158	165											
60%	114	119	124	129	134	140	146	152	158												
65%	114	119	124	129	134	140	146	152													
70%	114	119	124	129	134	140	146														
75%	114	119	124	129	134	140															
80%	114	119	124	129	134																
85%	114	119	124	129																	
90%	114	119	124																		
95%	114	119																			
100%	114																				

To this extent, the whole range of possible results for the year 2018 is depicted by figure 3.2.3.8 (see appendix B for 2016 and 2017). The weight attached to activities / events is reported in columns, the one for loans (physical and digital) in rows, while the importance of making facilities available for studying/ reading is the residual, given that they need to sum up to one.

The variability in the results is particularly high: the most positive estimate implies that productivity has almost tripled from 2015 to 2018 (+160%), while the most negative one reports an increase of around 14%. However, there are some considerations that still need to be made:

- *Productivity increased*. Although we do not know its magnitude, accepting the assumptions underlying the model, it can be safely affirmed that libraries have become more productive in the three years after 2015.
- *Activities steeply increased*. That is, as shown by the colours in figure 3.2.3.8, the higher the weight attached to the index representing activities, the higher the productivity increase.
- *Loans and visits for studying or reading increased at a slower and similar pace*. This can be explained by the low variability within the same column.

Furthermore, a more qualitative approach can be undertaken seeking to reduce the number of (potential) results. Firstly, although recent normative changes have reshaped the role of libraries, it is reasonable to believe that loans (physical and digital) are still the most important service they offer. It is difficult to believe that the introduction of activities in 2014/15 has already changed citizens' perception about the social role of public libraries, which continue to be seen as "the place where books and other materials can be borrowed". It follows that, even though activities represent a crucial service provided by libraries, they are probably not their main output.

Equal reasoning apply to the dimension related to self-studying. Approximatively translating these considerations into numbers allows to discard some of the estimations and focus on a restricted set, which is highlighted by black borders in the figure. Although the deviation in these numbers is still high, it is lower than when considering the whole table: the best scenario highlights a productivity increase of 35% from 2015 to 2018, while the worst one depicts an increase of 19% during the same period of time.

Finally, as anticipated above, the whole analysis hinges on the fact that inputs are assumed to be combined in the productive process according to their marginal productivity, which might not hold for volunteers. Indeed, it might be the case that volunteers are paid at a lower value than the value they add in the production of final outputs.

One way to proceed would entail the comparison of the hourly wage of volunteers and employees to infer the potential impact of an underpayment of the former. The question to answer is: "how would TFP change if volunteers were paid the same salary of employees?". However, as only the headcount of volunteers is available, their part-time factor cannot be precisely determined, so that the analysis needs to be based on assumptions. Firstly, the actual annual value of volunteers' inputs (AVC_i) is recalculated dividing the annual volunteers' compensation (VC_i) by an underpayment factor (UF), as explained by Equation (12).

$$AVC_i = \frac{VC_i}{UF} \quad (12)$$

Then, these "new" values are used instead of the real ones in the calculation of the aggregated annual input index. Therefore, the lower the underpayment factor UF (meaning that volunteers are payed way below their marginal productivity), the higher AVC_i and, in turn, the higher the share of volunteers in the total input index. Therefore, as the employment of these people is highly increasing over time, this implies that low UF translates into a high total inputs index, which in turn lower TFP.

Table 3.2.3.9 shows how the TFP index in 2018 would change according to a potential undervaluation of volunteers. For example, when UF is 0,5, they are paid only 50% of their marginal productivity. The lower and upper bounds represent the minimum and the maximum values of TFP within the area delimited by the qualitative assumptions proposed above (black borders in figure 3.2.3.8).

Table 3.2.3.9. The underpayment of volunteers and its effect on TFP

UF	%TFP (2018 vs. 2015)	
	Lower bound[29]	**Upper bound**[30]
1	+19%	+34%
0.5	+16%	+32%
0.2	+10%	+24%
0.1	+2%	+15%
0.08	-2%	+11%

As one would expect, TFP decreases as the underpayment factor decreases but it remains mainly positive. That is, only if volunteers are paid 8% of the value they add to the production, the lower bound of TFP is negative (-2%). Nevertheless, these are only some scenarios, which require further assessment. In fact, volunteers' marginal productivity can be better understood by analysing their tasks, activities and contributions to the production.

3.3 The Fire Department

The following analysis assesses productivity in the Dutch Fire Department, which is part of the broader area of the public services aimed at ensuring citizens' security and safety.

3.3.1 Framing the context

The Fire Department in The Netherlands (Brandweer) has a decentralized structure, embedded in a cooperative network in which various services and organizations work together in the field of public order and safety, fire services,

[29] w_L=80%; w_V=15%; w_A=5%
[30] w_L=65%; w_V=15%; w_A=20%

disaster and crisis management and medical assistance. Indeed, according to its website (Brandweer, 2020), the Dutch Fire Service is described as a partnership of all the 450 municipal brigades active on the national territory, which in turn are aggregated into 25 regional brigades, one for each safety region. This subdivision was introduced in 2010 to improve regional disaster and crisis management: in each of these regions, (i) the fire brigade, (ii) the police, (iii) the medical services, and (iv) the municipal services work together to coordinate and deliver more effective responses. Furthermore, the 25 regional chiefs form the Council of Fire Commanders (RBC), which has a steering and guiding role at the national level. It distinguishes five objectives to be pursued by every fire brigade in The Netherlands: (i) risk Management and Fire Safe Living, (ii) incident response, (iii) professional competence and knowledge, (iv) crisis management, (v) people and management.

The scattered organization of the Fire Department at a regional level reflects the diversity in the tasks and challenges that each fire brigade needs to cope with. That is, although there are common objectives (e.g. reducing response time, improving operational management, etc.), there are many regional differences due to geographical characteristics (rural areas vs. urban regions).

The disperse nature of the Dutch Fire Department is one of the reasons making considerations about the overall productivity level a complex task. There are no annual reports or data collection efforts at the central level. The second reason reflects the fact that fire brigades in The Netherlands work closely with other institutions and organizations in the so-called safety regions. In fact, some of these regions provide annual analyses, but the financial accountability of the fire services is aggregated with the other participant institutions' one (e.g. ambulances, police, et cetera). Therefore, data as wages and personnel volumes of the fire brigade alone are often not available in these publications and, consequently, the following analysis hinges mainly on data made public by CBS on StatLine.

3.3.2 Outputs and quality data

On StatLine data on (i) fire incidents, (ii) other type of incidents requiring the intervention of a fire brigade, and (iii) the average response time (the time between the notification of an incident and the moment when the fire department arrives on the scene) are available from 2013. Information about fire victims and injuries in The Netherlands has been retrieved from the website of the Instituut Fysieke Veiligheid, while data about real estate damages has been partially downloaded by the Verbond van Verzekeraars's online dashboard and partially sent directly from the institute to the author. Data on fire prevention and training are not available. Table 3.3.2.1 presents all the variables used in the following process of outputs calculation and aggregation.

Table 3.3.2.1. Output indicators

	2013	2019	Source
Fire incidents (nr.) FI_i	126 520	115 480	(CBS, 2020e)
Other incidents (nr.) OI_i	108 120	122 380	
Average response time (min.) RT_i	7.4	7.8	(CBS, 2020f)
Fire fatalities (nr.) V_i	33	28	(IFV, 2020)
Fire injuries (nr.) $IP_{s,i}$	172	202	
Average fire claim to insurers (euro) AC_i	2 420	3 440	(Verbond van Verzekeraars, 2020)

3.3.3 Process Analysis and Productivity Calculation

This evaluation of the productive process of fire brigades starts from their motto: “Minder brand, minder slachtoffers, minder schade” (fewer fires, fewer victims, less damages). That is, the ultimate objective of the Fire Department is to reduce the number of victims and economic damages due to fires and other natural catastrophes.

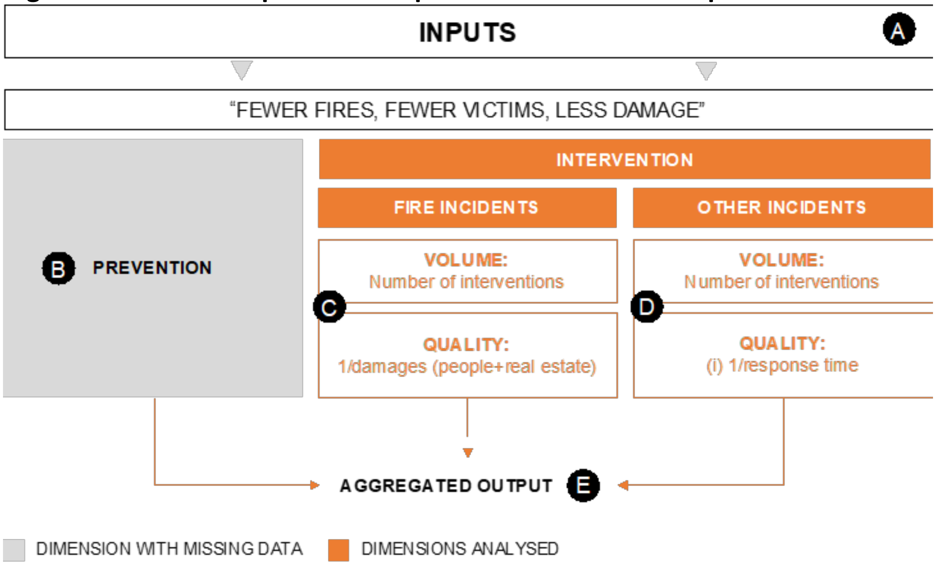
To measure the output of this institution and, consequently, calculate productivity, one would be tempted to look at the number of interventions adjusted by quality measures. Specifically, the higher the number of times that a fire brigade intervenes on an incident scene, the higher the output of the overall department. Nevertheless, understanding how inputs are combined to deliver services that seek to increase citizens’ safeness is a much more difficult matter, as fire brigades are active in two different dimensions: (i) prevention and (ii) intervention.

What makes this analysis complex is that prevention activities are aimed at reducing the number of fire events and, therefore, the number of interventions. In other words, simply focusing on the latter dimension, improvements in the final outcome valued by citizens (fewer fires) would be registered as a reduction in outputs and, consequently, productivity.

Bouckaert (1992) acknowledged the dual nature of the productive process of the Belgian fire department and built his analysis on the concepts of efficiency and efficacy. The first one is aimed at understanding the relationship between inputs and outputs, while the second one assesses how well outputs influence the final outcomes (effects, using his words). Although no calculations of aggregate productivity were provided, he operationalized his conceptual framework proposing some KPIs. Total output aggregated (i) the number of prevention activities, (ii) the number of fire suppression interventions, and (iii) the number of other emergencies that required a fire brigade to intervene. While final outcomes were (i) the ratio between the number of fires and inhabitants (to measure the effectiveness of prevention activities), (ii) casualties and economic damages per fire (to measure the effectiveness of fire repression).

Some other studies propose productivity calculations that hinge on the use of similar variables, Jadell (2002) represents an overview of these publications. The following data analysis for the Dutch fire services builds on these contributions, considering the fire department’s production as a dual-layered process, where prevention and interventions are combined with the final aim of reducing fire victims and damages. The main novelty of this study is indeed the direct connection between the economic impact of fires and the volume measures of fire brigades' outputs to obtain a final proxy of TFP. Figure 3.3.3.1. schematizes this approach, while the black letters highlight some issues that will be tackled on the following pages.

Figure 3.3.3.1. The productive process of the Fire Department



Issue A – Inputs

According to the COFOG classification, fire services expenses are recorded in the two-digit subsection 3.2 (Public order and safety), table 3.3.3.2 panel A summarizes these figures.

In the context of fire services, the main input-related issue regards how to treat information about volunteers, as they represent a very significant share of the total labour force (table 3.3.3.2 panel B). Unfortunately, data about firefighters in The Netherlands are available at a rather superficial level, as their participation in the productive process is recorded as a simple headcount with no information about total FTEs. Nevertheless, if considering paid employees to work full time (or supposing the ratio FTE/employee to be constant overtime) is a rather safe assumption, the same reasoning does not hold for volunteers; so that using simple headcount measures to infer about labour volume may lead to wrong results. Therefore, even if above it has been argued that a direct measure of labour is somehow preferable, the risk of heavily overlooking volunteers' contribution makes deflating the total current value of compensation of employees a safer approach.

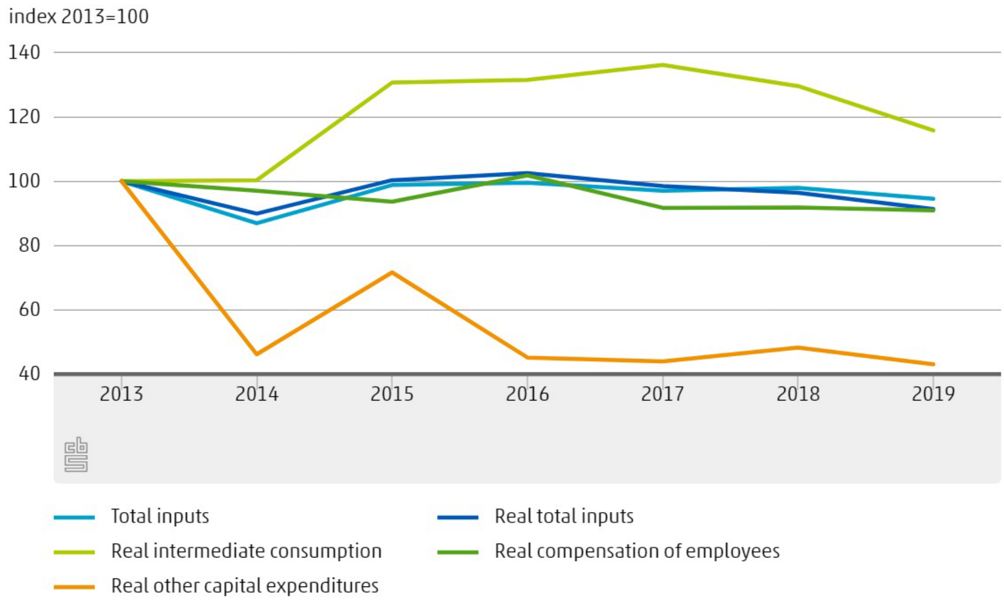
Note that, as the interest is in the evolution of the index over time, the only assumption that needs to be respected for this deflation to be reliable is that employees and volunteers' compensations change at the same pace. The same indirect deflation has not been done for libraries as more detailed data (FTEs for employees and disaggregated compensation) about the two dimensions were available.

Table 3.3.3.2. Inputs, deflators and volume measures of inputs for the Fire Department

	2013	2019	Average (2013/19)	Source
PANEL A				
Intermediate consumption (mln euro)	459	565	563	
Compensation of employees (mln euro)	914	926	913	(CBS,
Other current expenses (mln euro	31	24	24	2020a)
Capital expenditures (mln euro)	373	165	213	
PANEL B				
Total employees (headcount)	8 978	9 283	8 987	(CBS,
Total volunteers (headcount)	20 811	18 857	19 368	2019c)

The results of these deflations and their aggregation are reported in figure 3.3.3.3, where the dark blue line shows that the index reflecting the total volume of inputs decreased by 9% in the period 2013-2019. This reduction was triggered by a negative evolution in labour volume (-9%) and a drop in capital consumption (-57%), which accounted (on average), respectively, for 53% and 13% of the total fire department's expenditure.

Figure 3.3.3.3. Fire Department's inputs



Finally, one would estimate the impact of volunteers on the final TFP figures, as it may be the case that using cost shares for inputs aggregation is not an optimal method. Nevertheless, compensations of employees and volunteers are recorded at an aggregated level, so that considerations like the ones proposed for public libraries are not feasible.

Issue B – Fire Prevention

One of the two dimensions that need to be analysed to obtain information about the output of fire brigades relates to prevention activities. That is, an important function of the fire department aims to proactively reduce the number of incidents and their damages; this happens, for example, by carrying out periodic checks to smoke detectors and sprinkler systems in public facilities and private houses. Unfortunately, there is no information about this area, as the number of prevention interventions is not tracked by neither CBS nor the 'Brandweer'. Although the former fact forces to discard this dimension from the following TFP analysis, some considerations about its (potential) impact on the overall productivity are proposed at the end of this section.

Issue C – Fire Interventions
Fire interventions are at the core of this productivity analysis. As their volume measure (number of fires) is easily retrievable from StatLine (CBS, 2020e) the main challenge is to account for quality improvements. The adjustment proposed here discards the use of the average response time[31] (or the average time spent fighting a fire), considering that the final outcome valued by citizens and the policymaker is represented by (i) the impact on people's health (casualties and injures) and (ii) the economic damages caused on real estate properties. That is, casualties and fire damages are a function of the average response time and firefighters' skill level, among others. If this is true and data about the former effects are available, there is no need to consider other variables to adjust for quality. It is worth noting that approaches taking response time as adjustment can be misleading, as information about the shape of the function are usually overlooked. For example, it could be the case that a reduction in response time from 8 to 7 minutes does not have the same impact on damages as a reduction from 3 to 2 minutes.

Considering (i) real estate damages, (ii) casualties and (iii) injured people is not a novel approach (Jaldell, 2002); however, as described below, the following analysis attaches a monetary value to each of the three dimensions, reducing aggregation arbitrariness.

Firstly, as shown in equation (13), the annual value of real estate damages (RE_i) is constructed multiplying the number of fire incidents (FI_i) by the average claim (AC_i) reported by insurance companies for fire events (see table 3.3.3.1). It is noteworthy that the annual number of fire claims is lower than the number of interventions performed by fire brigades, as not all citizens / companies are insured against fire damages or qualify for a reimbursement. Therefore, the implicit assumption here is that the average claim to insurers is representative of the actual damage for both insured and non-insured events, which seems to be a reasonable premise.

$$RE_i = FI_i * AC_i \qquad (13)$$

Secondly, as shown in equation (14), the annual monetary impact of life losses (LL_i) in fire incidents is estimated according to the method of Statistical Life Value (VLS), which is increasingly common in the evaluation of policies that affect people's mortality risk.

[31] The average time between the emergency call and the arrive of the fire brigades on the scene.

$$LL_i = V_i * VLS \quad (14)$$

Where V_i is the annual number of fatal victims. The individual monetary value (VLS) is taken by a meta-analysis published by the OECD (Lindhjem, Navrud and Braathen, 2010), which represents a comprehensive review of the works available on this topic. The number[32], which is an average of all the estimation previously proposed by researchers, is adjusted for exchange rate and price levels and multiplied by the number of fatal victims in fire incidents in The Netherlands (table 3.3.3.1).

The third dimension entails the monetary losses caused by injured people, which can be summarized as (i) the cost of health treatment and hospitalization, (ii) the permanent reduction in the life's value of victims. Firstly, as reported in equation (15), the number of injured people ($IP_{s,i}$) is provided each year (i) at a disaggregated level according to the severity of the trauma (s): the first column of table 3.3.3.4 shows this categorization, the second one reporting the respective average share of victims between 2013 and 2019.

In order to estimate the annual healthcare costs of fire injuries (first part of equation (15)), the calculation builds on Ahn and Maitz (2012), who argue that the most important determinant of hospitalization costs for fire victims is the severity of the burn (TBSA[33]). They calculate these average costs for four categories, from the least acute lesion (TBSA 0%-9.9%) to the most serious one (TBSA +30%). As shown in table 3.3.3.4 (columns 3 and 4), this categorization is matched with the one proposed by the IFV for Dutch victims, so that the number of people in each category ($IP_{s,i}$) can be multiplied by the respective cost (IHC_s), leading to aggregated results that represent an annual measure of the total economic impact of fires on the health system.

Secondly, there is the need of taking long-term damages of fire injuries on people's lives into account. That is, if there are long-lasting effects of burns negatively affecting how people can conduct their daily life (e.g. partial inability), the total monetary measure of damage proposed above should account for these social losses. Table 3.3.3.4. (columns 5 and 6) attaches a value ranging from 0% to 15% of the VLS figure employed in the previous calculation to account for permanent injuries in each category ($\%PD_s$). This number increases with the severity of the burn and is linked with the magnitude of the healthcare cost. Note that, although this is a rather discretionary assumption (as permanent damage estimations are not available), these values remain constant over time, reducing the arbitrariness in the final aggregated result. Finally, healthcare costs and estimations of permanent damages are aggregated into annual total measures (I_i):

32 The VLS value is 6.256.000$ (in 2005 prices), which converted in euros and adjusted by inflation is 5 852 711 euro (in 2013 prices).

33 The severity is determined according to the percentage of body surface reporting burn injuries: %TBSA (Total Body Surface Area).

$$I_i = \sum_{s=1}^{S} IP_{s,i} * IHC_s + \sum_{s=1}^{S} IP_{s,i} * \%PD_s * VLS \quad (15)$$

Table 3.3.3.4. Injuries severity, healthcare cost and permanent damages

Burn Severity	Avg. % of total victims (2013-19)	TBSA%	Individual healthcare cost (IHC_s)	Permanent damage ($\%PD_s$)	Permanent damage
	%	%	2013 euro	% of VLS	2013 euro
No injuries	22		0	0	0
Slight smoke inhalation	43	0 - 9.9	16 050	1	58 527
Severe smoke inhalation	23	0 - 9.9	16 050	1	58 527
Minor burns	4	10 - 19.9	68 345	5	292 636
Severe burns	4	+30	203 482	15	877 907
Other injury	5				
	(IFV, 2020)	(Ahn and Maitz, 2012)		author's assumptions	

Furthermore, in order to sterilize the effect of price changes on these figures, price corrections[34] are applied before calculating the annual measure of average fire damages (DMG_i), as described by equation (16).

$$DMG_i = \frac{RE_i + LL_i + I_i}{FI_i} \quad (16)$$

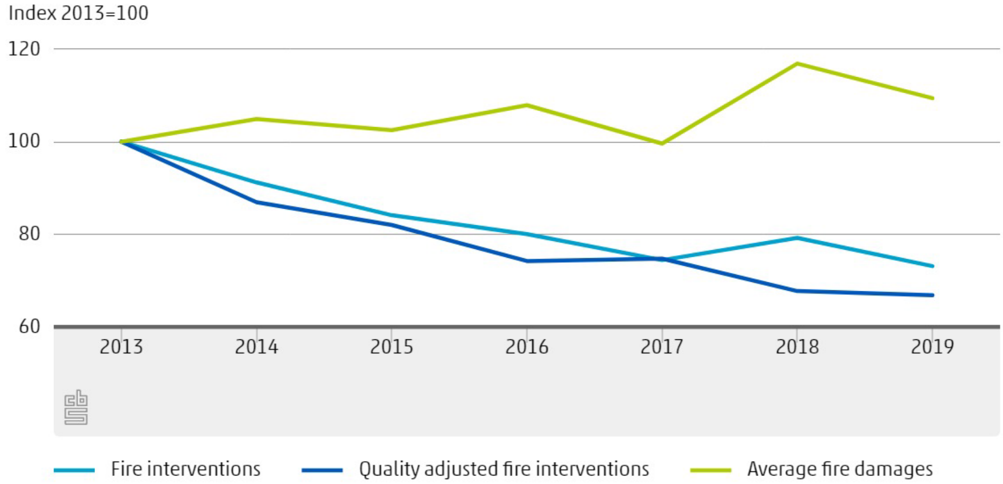
Finally, according to equation (17), the index representing the volume of fire incidents (FI_i) is divided by the one describing the annual average monetary damage (DMG_i) to obtain a quality-adjusted measure of fire interventions (YF_i). The trend is reported in figure 3.3.3.5.

$$YF_i = \frac{FI_i}{DMG_i} \quad (17)$$

In fact, the average damage caused by fire incidents increased by 9% between 2013 and 2019, while their volume was decreasing by 33%. This, in turn, made the adjusted output plunge even further (-37%).

[34] The CPI index (CBS, 2020c) is used as a deflator for "people damages" ($LL_i and\ I_i$), while an index representing the development of the price of (new) real estate properties (CBS, 2020g) is employed to deflate the related damages. This choice reflects the fact that costs of rebuilding houses increase at a different rate than the overall price level, so that a price index for newly built houses is more appropriate. As house prices increased more than the overall price level, the final estimation of TFP is lower than the case in which CPI index is used to deflate all the three dimensions. Nevertheless, this difference would be rather small: productivity (first row of table 3.3.3.7) would only decrease by 2% in 2019.

Figure 3.3.3.5. Fire interventions, damages and quality-adjusted indices



The intuition to understand how this "monetary correction" (DMG_i) works can be made clear by the following example: if the volume of fire interventions remains constant over time but the average damage increases, then productivity will decrease ceteris paribus, as, presumably, firefighters perform their tasks less effectively.

Note that since the three different dimensions are expressed in monetary terms, there are no implicit trade-offs embedded in the aggregation. Assume there is an increase in one of the dimensions (real estate damages) at the expense of a second one (injuries): will the overall quality index value decrease or increase? If we do not explicitly translate volumes into monetary terms, the answer to this question is given by the marginal rate of substitution, which can be interpreted as the ratio between the weights attached to the two dimensions. Therefore, if the assumptions behind the aggregation method are arbitrary and the potential trade-offs are not fully made clear and treated, the quality index may lead to perverse valuations.

Nevertheless, before turning to the next issue, it is worth noting that attaching monetary valuations to lives, healthcare processes, and long-lasting fire injuries is a complex and multifaceted task. The calculation proposed above makes rather arbitrary assumptions, as data are not available at a detailed level. Consequently, potential improvements entail the use of data coming directly from hospitals and health facilities, which could give specific information about each single patient. However, since the interest is in the change over time, the importance and impact of the variability in the absolute values are reduced. Finally, another caveat to this calculation is the fact that the numbers of casualties and injured people used in the analysis and reported in table 3.3.2.1 are presumably underestimated, as they only include incidents in residential contexts (e.g. car fires are not accounted for).

Issue D – Other Interventions

Following the framework proposed in Figure 3.3.3.1, interventions are sub-divided in fire-related ones and others. As the volume of both dimensions can be measured by the quantity of incidents, one would apply to other interventions the same monetary-based approach to account for quality changes. Nevertheless, attaching an economic value to these incidents is a more complicated matter, given their multifaceted nature and the lack of precise data. For this reason, the quality-adjusted index (YO_i) is calculated dividing the volume indicator (OI_i) by the one representing the inverse of the average response time (RT_i) of fire brigades (see Table 3.3.2.1 and equation (18)).

$$YO_i = \frac{OI_i}{RT_i} \tag{18}$$

Although the supposition is that lower response times translate into less damage, there is no information about the actual shape of this relation. That is, two assumptions are made: (i) a monotonic relationship between damages and response time and (ii) implicit marginal rates of substitution between them, deriving from the equal weights attached to these variables. As no other data can be used to infer about this matter, further research is needed to better consider the impact of these assumptions on the quality-adjusted output.

Issue E– Aggregation and TFP Calculations
All the information required to describe the intervention-related side of the fire department's productive process has been explained and now, needs to be aggregated into a single output measure (Y_i).

$$Y_i = wF_i * YF_i + wO_i * YO_i \tag{19}$$

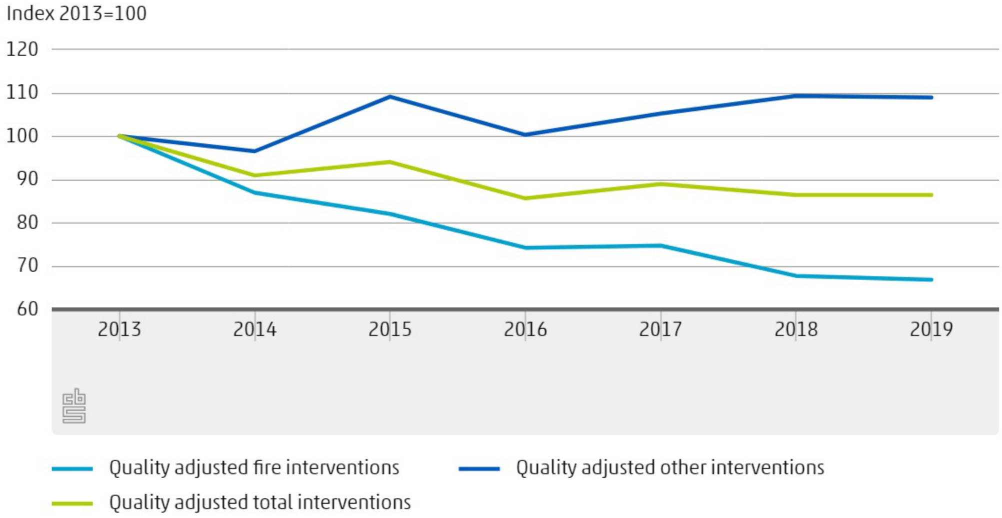
The assumption made here is that fire and other interventions have the same value to final consumers (citizens)[35]. Consequently, as shown in equation (19), the two dimensions' volumes are used as weights for the combination of the two indices, so that $wF_i = \frac{FI_i}{FI_i+OI_i}$ and $wO_i = \frac{OI_i}{FI_i+OI_i}$. The trends are shown in figure 3.3.3.6.

Input and output measures can now be employed in the calculation of productivity, which, as explained above, will not consider one of the two fundamental functions of fire services: prevention. This productivity calculation is reported in the first row of Table 3.3.3.7 (i) and shows that TFP decreased by 5% between 2013 and 2019, meaning a negative average annual growth rate of -0.92%.

It is worth highlighting that the year 2016 displays a stronger drop in productivity than the overall trend. This plunge is (mainly) triggered by a reduction in the quality adjustment of fire interventions, which, in turn, decreased due to a peak in the annual number of victims in 2016. Note that, fortunately, the annual number of fatalities in fire events is a small value to which, however, a high economic impact is attached (VSL), so that even small changes can provoke significant variation in the final productivity measure. To reduce the impact of such variability, a possible strategy may entail the use of moving averages.

[35] Note that this can be considered an arbitrary assumption, as it may well be the case that some interventions need a higher weight than others (fire in a house vs. cat up in a tree). However, the granularity in the data does not allow to take this into account.

Figure 3.3.3.6. Fire brigades'output



The second row (ii) of table 3.3.3.7 shows the TFP index when the number of fatalities employed in the calculation is the average between the actual figure and the ones of the two years before. As it can be noted, the numbers now show a smoother trend than the one presented before. Nevertheless, to overcome the lack of data regarding prevention, table 3.3.3.7 proposes three hypothetical scenarios to understand the potential impact of variation in this dimension on the final productivity of fire services: (i) decreasing quality-adjusted prevention activities output (-2% per year), (ii) stable number of inspections, and (iii) increasing adjusted prevention interventions (+2% per year).

The two dimensions (interventions and preventions) are aggregated into a total output measure (TY_i) according to equation (20), where wI and wP are the respective weights, while P_i is the quality-adjusted volume measure of prevention activities.

$$TY_i = wI * Y_i + wP * P_i \tag{20}$$

For each of these scenarios, as the weights to aggregate prevention and interventions (wI and wP) are unknown, a simplified sensitivity analysis is put forward, where the weight attached to the former dimension is reported in the table and the one for the latter is calculated as the residual.

Assessing the figures in table 3.3.3.7, it can be seen that the variability in the results of the different sensitivity analyses is relatively high. In fact, excluding the rather unlikely situation in which prevention activities make up 75% of the total output of fire brigade, TFP registered a change ranging between -5% and +8% (between 2013 and 2019), translating into an -0.8%/+1.3% average annual growth rate.

Tabel 3.3.3.7. Fire Department TFP and Sensitivity Analysis

		TFP (2013=100)						
		2013	2014	2015	2016	2017	2018	2019
(i)	Base scenario: only interventions							
		100	101	94	83	90	90	95
(ii)	Base scenario: only interventions (MA)							
		100	100	93	87	87	90	90
(iii)	Scenario 1: (prevention -2%/y.)							
	wP: 25%	100	103	94	85	91	91	95
	wP: 50%	100	105	95	88	92	92	96
	wP: 75%	100	107	95	90	93	93	96
(iv)	Scenario 2: (prevention +0%/y.)							
	wP: 25%	100	103	95	87	93	93	98
	wP: 50%	100	106	97	90	96	96	102
	wP: 75%	100	109	98	94	99	100	106
(v)	Scenario 3 (prevention +2%)							
	wP: 25%	100	104	96	88	95	95	101
	wP: 50%	100	107	99	93	100	101	108
	wP: 75%	100	110	101	98	105	108	115

In conclusion, given the caveats of the computation process highlighted above and the data available, it is difficult to safely determine in which direction TFP is moving. Nevertheless, recall that it has been explained above how prevention activities exert two opposite pressures on the final output, as increases in prevention are aimed at reducing the number of fires (and, therefore, interventions), which, in turn, will be registered as a decrease in the output of the respective dimension. Moreover, as shown in 3.3.3.7, note that the sharp decrease in the quality-adjusted output of the intervention-related dimension is determined by a constant decrease of the volume of fire interventions, read fewer fires year after year.

Putting together these two pieces of information, one could argue that one reason for the number of fires to decrease over time may be an increasing effort in prevention. This would make the third scenario of 3.3.3.7 the most likable, meaning productivity in the fire department slightly increased in the years 2013-2019; however, more and better data are needed to understand how much this annual growth differ from zero.

4. Conclusion

The methodological and practical discussions above have highlighted the main issues arising when one analyses productivity in the public sector. To briefly summarize what was tackled on the previous pages, the lack of meaningful prices and transactions with final consumers makes public institutions' TFP calculations a more complex matter than for the private sectors. In fact, while for the latter volume and value information is readily available or easily computable, the same does not hold for the public sector. It has been shown how volume output indices need to be manually constructed, following a careful analysis of the institution-specific production process. This has been done for public libraries, where the final output is explained through the aggregation of loans, activities and visits information, and the fire department, which aims to increase citizens' safety by combining prevention activities and emergency interventions. Furthermore, these case studies emphasize the importance of direct quality adjustments, as final consumers value not only the quantity of a product (or service) but also its quality.

Besides the direct institution-specific policy implications of the TFP numbers, a broader conclusion can be drawn: the output=input method inadequately seizes the value of the public sector's contribution to the total economy. In fact, this method considers TFP to be always zero, even though the two case studies clearly show a different pattern. The public libraries demonstrate a clear and sharp increase in productivity, while safe conclusions for the fire department are difficult to derive. In any case, if the TFP figures are even slightly positive, one would consider the whole national productivity level reported by national statistics to be underestimated. Consequently, if productivity is higher than zero and the public sector's production is accounted for summing up inputs employed in the productive process, economic growth is also undervalued. Note that this has massive implications, especially for developed countries such as The Netherlands, where the State has a pivotal role in the economy.

Furthermore, one of the main objectives of this thesis was to analyse the extent to which productivity in public institutions can be accurately calculated. It is indeed clear that final TFP numbers for public institutions depend to a greater degree on the measurement choices that are taken in the analytical process than for private organizations, emphasizing statisticians' role. As demonstrated in the previous case studies, these decisions play a crucial role in determining the variability of the final figures, so that additional importance needs to be put in initial measurement choices.

One significant example is the contribution of volunteers to the production, which has been assumed to be represented by the amount of money they receive as compensation. Nevertheless, it may well be the case that these figures do not proxy their actual contribution, so that the obtained TFP indices may be overestimated. This does not mean that productivity in the public sector cannot be calculated, but that these caveats must be acknowledged, and further effort needs to be employed in improving the measurement framework and the related available data.

In particular, it is worth highlighting the crucial role of institution-specific cost information, so much that its absence has made necessary the use of sensitivity analyses for the aggregation of output indices. Indeed, under the assumption that public institutions are steered to meet consumers' preferences, their cost structures are representative of the latter's desires (see 2.5). Furthermore, the previous analysis highlights the importance of other improvements in the data collection process, which culminated in the following recommendations.

Firstly, on the inputs side, the main potential enhancements regard labour computations. Quality in the labour force (skill level) needs to be accounted for and the availability of salary information could allow for better adjustments in volume computations. Secondly, when it comes to the output side of the calculation, some specific issues need to be tackled; starting from libraries, there are two main recommendations:

- Deepening the analysis on how consumers value collection variety to better understand how quality adjustments to loans volume measures should be applied.
- Improving the detail in the data reporting the number of visits to have separate measures for people using the facilities to study/ read, to borrow books and other materials, or to participate activities.

Additionally, the fire department requires:

- Including information on prevention activities, as the consideration of this dimension in the calculation of the whole institution's productivity is pivotal.
- Incorporating damages related to fire incidents (casualties, injuries, and real estate) in other contexts other than the residential one. Moreover, better information about the severity of people fire injuries, perhaps through the use of hospitals' database, may help to improve their conversion to monetary terms.
- Including data about the disaggregated nature of other interventions to allow similar monetary analyses as the ones carried out for fire events.

Finally, the availability of detailed data about all the single departments active within the same institution (e.g. inputs and outputs for each single library in The Netherlands) would allow the modelling and the regression of its production function, to understand to what extent additional output stems from increasing productivity.

Literature

Ahn, C. S. and Maitz, P. K. (2012). The true cost of burn. Burns, 38, 967-974.
Atkinson, A. B. (2005). The Atkinson review: final report. Measurement of government output and productivity for the national accounts. Palgrave Macmillan.

Basu, S. and Fernald, J. G. (2002). Aggregate productivity and aggregate technology. European Economic Review, 46(6), 963-991.

Basu, S. and Fernald, J. G. (2002). Aggregate productivity and aggregate technology. European Economic Review, 963-911.

Basu, S., Inklaar, R. and Wang, C. J. (2011). he value of risk: measuring the service output of US commercial banks. Economic Inquiry, 49(1), 226-245.

Basu, S., Pascali, L., Schiantarelli, F. and Serven, L. (2012). Productivity and the Welfare of Nations. NBER WORKING PAPER SERIES https://www.nber.org/papers/w17971.

Bibliotheekinzicht. (2020, October). Dashboard Bibliotheekstatistiek (Gegevenslevering Wsob). Retrieved from bibliotheekinzicht.nl: https://www.bibliotheekinzicht.nl/instrumenten/databank/dashboard-bibliotheekstatistiek-2018

Blank, J. L. T. (2018). Measuring the performance of local administrative public services. Business Research Quarterly (21), p. 251-261.

Blank, J.L.T. and Van Heezik, A.S. (2019). Productiviteit en allocatie op het terrein van veiligheid en justitie, 1980-2016; een essay. Delft: IPSE Studies

Blank, J. L. T. and Van Heezik, A.S. (2020). Policy reforms and productivity change in the judiciary system: a cost function approach applied to time series of the Dutch judiciary system between 1980 and 2016. International Transactions in Operational Research (27), p. 2002-2020

Blank, J.L.T. and Van Heezik, A.S. (2021). Productiviteit van overheidsbeleid; Samenvatting productiviteitstrends in de Nederlandse justitie en veiligheid, 1980-2018. Delft: IPSE Studies

Bouckaert, G. (1992). Productivity analysis in the public sector: The case of the fire service. International Review of Administrative Sciences, 58(2), 175-200.

Brandweer. (2020, November 14). De regio's. Retrieved from Brandweer.nl: https://www.brandweer.nl/ons-werk/de-regio-s

Brynjolfsson, E., Hu, Y. and Smith, M. D. (2003). Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. Management Science, 49(11), 1580-1596.

Caselli, F. (2005). Accounting for Cross-Country Income Differences. Handbook of Economic Growth, 679-741.

CBS. (2019a, June 26). Overheid; overheidsproductie en -consumptie 1995-2018 . Retrieved from StatLine: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82568NED/table?dl=42408

CBS. (2019b, November 21). Openbare bibliotheken. Retrieved from StatLine: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70763ned/table?ts=1603721669403

CBS. (2019c, June 20). StatLine. Retrieved from Brandweerpersoneel; beroeps en vrijwilligers, rangniveau, opleiding, regio: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/71482NED/table?fromstatweb

CBS. (2019c, November 21). Landelijke digitale openbare bibliotheek; gebruik, accounts en licenties. Retrieved from StatLine: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83804NED/table?ts=1603721803349

CBS. (2020a, September 30). Overheidsuitgaven en bestedingen; functies, transacties, overheidssectoren Gewijzigd op: 30 september 2020. Retrieved from StatLine: https://opendata.cbs.nl/statline#/CBS/nl/dataset/84122NED/table?ts=1603796683786

CBS. (2020b, October 27). Prijs van arbeid; index 2015=100; nationale rekeningen. Retrieved from StatLine: https://opendata.cbs.nl/#/CBS/nl/dataset/84183NED/table?dl=4464D

CBS. (2020c, October 8). Jaarmutatie consumentenprijsindex; vanaf 1963. Retrieved from StatLine: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70936ned/table?ts=1603819244855

CBS. (2020d, June 24). Bruto investeringen in vaste activa; bestemming, nationale rekeningen. Retrieved from StatLine: https://opendata.cbs.nl/#/CBS/nl/dataset/84090NED/table?dl=44645andts=1604832938084

CBS. (2020e, March 23). Branden en hulpverleningen; alarmering van de brandweer, regio . Retrieved from StatLine: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83122NED/table?ts=1606055212684

CBS. (2020f, March 20). Branden; reactietijden van de brandweer, regio . Retrieved from StatLine: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83123NED/table?ts=1605369071158

CBS. (2020g, September 29). Nieuwbouwwoningen; outputprijsindex bouwkosten, 2015=100 . Retrieved from StatLine: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83908NED/table?ts=1608625456262

Comin, D. (2010). Total Factor Productivity. In S. Durlauf and L. Blume, Economic Growth (pp. 260-263). London: Palgrave Macmillan.
Coyle, D. (2015). GDP: a brief but affectionate history-revised and expanded edition. Princeton University Press.

CPB (2017). Productiviteitsontwikkeling van de Nederlandse overheid. The Hague: CPB.

CPB (2018). Productiviteitsgroei Nederlandse overheid 2012-2015; Twee methoden vergeleken. The Hague: CPB.

Cutler, D. M., Ghosh, K., Messer, K., Raghunathan, T., Rosen, A. B. and Stewart, S. T. (2020). A Satellite Account for Health in the United States . National Bureau of Economic Research.

Cutler, D., McClellan, M., Newhouse, J. and Remler, D. (1998). Are medical prices declining? Evidence from heart attack treatments. The Quarterly Journal of Economics, 113(4), 991–1024.

Diewert, E. W. (1976). Exact and Superlative Index Numbers. Journal of Econometrics, 115-145.

Diewert, E. W. (2011). Measuring productivity in the public sector: some conceptual problems. Journal of Productivity Analysis, 36(2), 177-191.

Diewert, E. W. (2017). Productivity Measurement in the Public Sector: Theory and Practice. Vancouver School of Economics.

Diewert, E. W. and Nakamura, A. O. (2005). Concepts and Measures of Productivity: An Introduction. In Lipsey and Nakamura, Services Industries and the Knowledge Based Economy (pp. 19-38). University of Calgary Press.

Douglas, J. (2006). Measurement of Public Sector Output and Productivity. New Zealand Treasury.

Dunleavy, P. (2017). Public sector productivity: measurement challenges, performance information and prospects for improvement. OECD Journal on Budgeting, 17(1), 1-28.

Dunleavy, P. and Hood, C. (1994). From old public administration to new public management. Public money and management, 14(3), 9-16.

Englander, S. A. and Gurney , A. (1994). Medium-term determinants of OECD productivity. OECD Economic Studies, 22(1), 49-109.

Eurostat. (2001). Handbook on price and volume measures in national accounts.

Eurostat. (2010). European system of accounts. Luxembourg: European Union.

Eurostat. (2016). Handbook on prices and volume measures in national accounts. Publications Office of the European Union.

Fernald, J. and Inklaar, R. (2020). Does Disappointing European Productivity Growth Reflect a Slowing Trend? Weighing the Evidence and Assessing the Future. International Productivity Monitor, 104-135.

Gruen, D. (2012). The Importance of Productivity. Productivity Perspectives Conference. Productivity Commission (Australian Bureau of Statistics).

Heathfield, D. F. and Wibe, S. (1987). An Introduction to cost and production function. London: MacMillan Education.

Helpman, E., Melitz, M. J. and Yeaple, S. R. (2004). Export versus FDI with heterogeneous firms. American Economic Review, 94(1), 300-316.
Hill, P. T. (1977). On Goods and Services. Review of Income and Wealth, 23(4), 315-338.

HM Government. (1998). UK National Accounts. Blue Book.

Hsieh, C.-T. and Klenow, P. J. (2010). Development Accounting. American Economic Journal, 207-223.

Hulten, C. R. (2001). Total Factor Productivity. A Short Biography. In C. R. Hulten, E. Dean and M. J. Ha, New Developments in Productivity Analysis (pp. 1-54). University of Chicago Press.

Hulten, C. R. (2010). Growth accounting. Handbook of the Economics of Innovation, 987-1031.

Huysmans, F. and Hillebrink, C. (2008). The future of the Dutch public library: ten years on. The Hague: Netherlands Institute for Social Research (SCP).

IFV. (2020, October). Kerncijfers Veiligheidsregio's. Retrieved from Kerncijfers Veiligheidsregio's: https://kerncijfers.ifv.nl/dashboard/kerncijfers-veiligheidsregio-s/

Jaldell, H. (2002). Essays on the performance of fire and rescue services. Gothenburg University Publications, n.116.

Jorgenson, D. W. and Fraumeni, B. (1989). The Accumulation of Human and Nonhuman Capital, 1948–84. In R. E. Lipsey and H. S. Tice, The Measurement of Saving, Investment and Wealth (Vol. 52). University of Chicago Press.

Kaufman, H. (1976). Are government organizations immortal? Brookings Institute Press.

Kwink Groep. (2019). Evaluatie Wet stelsel openbare bibliotheekvoorzieningen. The Hague: Kwink Groep.

Lau, E., Lonti, Z. and Schultz, R. (2017). Challenges in the measurement of public sector productivity in OECD countries. International Productivity Monitor, 180-195.

Levitt, M. S. and Joyce, M. S. (1987). The Growth and Efficiency of Public Spending. Cambridge: Cambridge University Press.

Lindhjem, H., Navrud, S. and Braathen, N. A. (2010). Valuing lives saved from environmental, transport and health policies: a meta-analysis of stated preference studies. Paris: OECD.

Lipsey, R. G. and Carlaw, K. (2000). What does total factor productivity measure? International Productivity Monitor, 1(31), 31-40.

Mellens, M., Deijl, C. and Van Veldhuizen (2020) Kan Nederlandse overheid Beaumol verslaan? The Hague: Centraal Planbureau.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica, 71(6), 1695–1725.

Moolenaar, D.E.G. (2019) Productiviteitsontwikkeling in het Justitieveld; een verkenning van mogelijkheden. Cahier 2019-5.

Niaounakis, T. and Van Heezik, A.S. (2020) Productiviteit Kadaster; Een analyse van de productiviteitsontwikkeling bij het Kadaster tussen 2015 en 2019. Delft: IPSE Studies.

Niaounakis, T. Van Heezik, A.S. and Blank, J.L.T. (2020) Productiviteit uitvoering sociale zekerheid; Een analyse van de productiviteitsontwikkeling bij UWV en SVB tussen 2002 en 2018. Delf: IPSE Studies.

OECD. (2017). Classification of the Functions of Government (COFOG). Government at a Glance 2017.

Oulton, N. (2016). The Mystery of TFP. International Productivity Monitor(31), 68-87.

Pritchard, A. (2002). Measuring productivity change in the provision of public services. Economic Trend, 20-32.

Productivity Commission. (2017). Non-market Sector Productivity, Shifting the Dial: 5 year Productivity Review. Canberra.

Raad voor Cultuur. (1998). Advies over de bestuurlijke organisatie van het openbare bibliotheekwerk. The Hague: Raad voor Cultuur.

Rainey, H. G. (2009). Understanding and Managing Public Organizations. Jossey-Bass. A Wiley Imprint.

Rainey, H. G. and Chun, Y. H. (2005). Public and Private Management Compared. The Oxford Handbook of Public Management.

Rosenberg Hansen, J. and Ferlie, E. (2016). Applying strategic management theories in public sector organizations: developing a typology. Public Management Review, 18(1), 1-19.

Schreyer, P. (2010). Towards Measuring the Volume Output of Education and Health Services. A Handbook. OECD Statistics Working Papers, No. 2010/02.

Schreyer, P. (2012). Output, outcome, and quality adjustment in measuring health and education services. Review of Income and Wealth, 257-278.

Schreyer, P. and Pilat, D. (2001). Measuring productivity. OECD Economic Studies, 33(2), 127-170.

Simpson, H. (2009). Productivity in public services. Journal of Economic Surveys, 23(2), 250-276.

Solow, R. M. (1956). A contribution to the theory of economic growth. The quarterly journal of economics, 70(1), 65-94.

Stiglitz, J. E., Sen, A. and Fitoussi, J.-P. (2009). Report by the Commission on the Measurement of Economic Performance and Social Progress.

Syverson, C. (2011). What determines productivity. Journal of Economic literature, 49(2), 326-365.

Triplett, J. (2006). Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes. OECD Publishing.

Verbond van Verzekeraars. (2020, October). Woningbranden. Retrieved from Woningbranden: https://bi.verzekeraars.nl/db/woningbranden2019.html

Vereniging van Openbare Bibliotheken. (2005). Richtlijn voor basisbibliotheken. Retrieved from debibliotheken.nl: https://www.debibliotheken.nl/wp-content/uploads/2019/10/2005_richtlijn-voor-basisbibliotheken_vob_vng.pdf

Wallace, R. J. (1977). Productivity Measurement in the Fire Service. Public Productivity Review, 2(3), 12-36.

Appendix A – Issues on the aggregation of inputs

Works on Government productivity have often focused on the output side, overlooking the importance of identifying, deflating, and allocating the right inputs to each activity. However, it is worth briefly tackling three issues relative to the denominator of productivity calculations: (ii) capital input computation, and (iii) labour input computation.

Capital Input Computation

According to Diewert (2011), the procedures to compute input volumes are fairly satisfactory and lead to solid results; however, there is still an amount of discussion with regards capital services measures. The System of National Account (2009) calculates capital services in the public sector only by depreciation (wear and tear), whereas for the private market, the calculation method includes the opportunity cost of capital[36] as well. This leads to an underestimate of governments costs (and hence GDP) and it is in open conflict with the fundamental principles of a common standard between market and non-market sectors. The Atkinson Review (2005, page 49) insists on the importance of including the opportunity cost of the capital in the total calculation of input by "*adding to the capital consumption an interest charge on the entire owned capital*".

Diewert (2017) discusses this issue and proposes a general formula to calculate the annual net cost of using the services of capital inputs, which is:

$$u^t = [r^t - i^t + \delta(1 + i^t)]\, P^t \qquad (1)$$

Where u^t is the total adjusted capital cost, r^t represent the one-period financial opportunity cost, δ is the depreciation rate, i^t is the inflation rate, and P^t is the purchase price of the good.

In the absence of consideration about the opportunity cost of capital (r^t); the input figure will be based only on the second part of the formula, therefore, being underestimated.

Labour Input Computation

The second cost item requiring particular attention is labour, not just because it has been increasingly employed in government production over the last decades, but also because it can be measured by both a direct (counting and weighting the number of hours worked) and indirect methods (deflating total expenditure by a price index), which should lead to equivalent results. The direct method is usually more employed than the indirect one, but just simply counting the number of employees and computing it as a volume measure of labour will not lead to a

36 Revenue coming from the capital good had it been employed for other uses (rented).

reliable index, since it fails to record workforce's quality; so that some refinements are needed.

Firstly, according to Schreyer and Pilat (2001), the simple employees' headcount should be converted in the total number of hours worked. This is important for two main reasons: (i) it would mirror a possible shift in the number of hours worked per person[37]; (ii) it would account for a more flexible labour market, reflecting peaks and dips in labour demand, which are increasingly frequent in developed economies. The authors discourage the use of full-time equivalent jobs (FTE), since it does not adjust for changes in the average number of hours worked by full-time employees (e.g. new legislations or collective agreements), and its comparability across countries is limited because of a lack of international methodological standards.

Secondly, and most importantly, our measure of labour inputs should gauge two characteristics of the workforce: time spent and skill level. Counting the total number of hours worked takes care of the first aspect, while it leaves the latter not accounted for. Therefore, to fully understand the evolution of skills in the workforce, changes in the quality of labour should be considered (Atkinson, 2005).

Schreyer and Pilat (2001) suggests a method based on the theory of the firm[38]: assuming that a firm will hire new employees to the point where their cost is equal to the additional revenues they can generate, it follows that wages can be used to obtain relative weights to aggregate 'different qualities' of hours worked.

To calculate the evolution of the number of quality-adjusted hours worked, we should start by computing the year-on-year change in the total number of (unadjusted) hours worked ($\Delta logHW_{i,t}$), as represented in equation (2)). Note that t represents the temporal dimension and i, the skill-based categories.

$$\Delta logHW_{i,t} = logHW_{i,t} - logHW_{i,t-1} \qquad (2)$$

Then, according to equation (2), the shift in the relative hourly wage ($\overline{w}_{i,t}$) should be computed for every skill category. Note that $w_{i,t}$ represents the relative wage for each category, which is calculated as a share of the sum of all hourly wages for every category. After computing the relative wages, the simple average of two consecutive periods is calculated, as shown in equation (3).

$$\overline{w}_{i,t} = \frac{1}{2}(w_{i,t} + w_{i,t-1}) \qquad (3)$$

Finally, the number of quality-adjusted hours worked ($\Delta logAHW$) is calculated as the weighted change in the total number of hours worked ($\Delta logHW_{i,t}$), where the weight is $\overline{w}_{i,t}$, as shown in equation (4).

37 Such shifts may be due to a move towards more paid vacations, shorter "normal" hours for full-time workers and greater use of part-time work.
38 "*Labour of a certain type will be hired up to the point where the cost of an additional hour of labour is just equal to the additional revenue that using this labour generates*" (Schreyer and Pilat, 2001).

$$\Delta AHW = \sum_{i=1}^{n} \overline{w}_{i,t} * \Delta logHW_{i,t} \qquad (4)$$

The higher the granularity in the data, the more every skill category can be broken down into different sub-categories (occupations) and, if the number of hours worked and the relative wage are separately recorded for each employee, the maximum precision can be reached (in this case, the subscription i would represent the single worker). However, as noted by Schreyer and Pilat (2001), such information requirements are highly demanding and datasets about hours worked and wages are costly and rarely available.

Literature

Atkinson, A. B. (2005). The Atkinson review: final report. Measurement of government output and productivity for the national accounts. Palgrave Macmillan.

Diewert, E. W. (2011). Measuring productivity in the public sector: some conceptual problems. Journal of Productivity Analysis, 36(2), 177-191.

Diewert, E. W. (2017). Productivity Measurement in the Public Sector: Theory and Practice. Vancouver School of Economics.

Schreyer, P. and Pilat, D. (2001). Measuring productivity. OECD Economic Studies, 33(2), 127-170.

United Nations, European Commission, International Monetary Fund, OECD and World Bank. (2009). System of national accounts 2008. United Nations

Appendix B – Libraries' TFP, 2016 and 2017

Figure B.1. Libraries' TFP 2016 (2015 = 100)

Loans \ Activities	0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%
0%	109	110	110	111	112	113	113	114	115	116	116	117	118	119	120	120	121	122	123	124	125
5%	108	109	110	111	111	112	113	114	114	115	116	117	117	118	119	120	121	121	122	123	
10%	108	109	109	110	111	111	112	113	114	115	115	116	117	118	118	119	120	121	122		
15%	107	108	109	109	110	111	112	112	113	114	115	115	116	117	118	119	119	120			
20%	107	107	108	109	110	110	111	112	113	113	114	115	116	116	117	118	119				
25%	106	107	108	108	109	110	111	111	112	113	114	114	115	116	117	117					
30%	106	106	107	108	108	109	110	111	111	112	113	114	114	115	116						
35%	105	106	106	107	108	109	109	110	111	112	112	113	114	115							
40%	105	105	106	107	107	108	109	110	110	111	112	113	113								
45%	104	105	105	106	107	108	108	109	110	110	111	112									
50%	103	104	105	106	106	107	108	108	109	110	111										
55%	103	104	104	105	106	106	107	108	109	109											
60%	102	103	104	104	105	106	107	107	108												
65%	102	103	103	104	105	105	106	107													
70%	101	102	103	103	104	105	105														
75%	101	101	102	103	104	104															
80%	100	101	102	102	103																
85%	100	100	101	102																	
90%	99	100	101																		
95%	99	99																			
100%	98																				

Figure B.2. Libraries' TFP 2017 (2015 = 100)

Loans \ Activities	0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%
0%	109	112	115	118	121	125	128	132	135	139	143	147	151	155	159	163	168	172	177	182	187
5%	109	112	115	118	122	125	128	132	135	139	143	147	151	155	159	164	168	173	177	182	
10%	109	112	115	119	122	125	129	132	136	139	143	147	151	155	160	164	168	173	178		
15%	110	113	116	119	122	125	129	132	136	140	143	147	151	156	160	164	169	173			
20%	110	113	116	119	122	126	129	133	136	140	144	148	152	156	160	164	169				
25%	110	113	116	119	122	126	129	133	136	140	144	148	152	156	160	165					
30%	110	113	116	119	123	126	129	133	137	140	144	148	152	156	161						
35%	110	113	116	120	123	126	130	133	137	141	144	148	152	157							
40%	111	114	117	120	123	126	130	133	137	141	145	149	153								
45%	111	114	117	120	123	127	130	134	137	141	145	149									
50%	111	114	117	120	124	127	130	134	138	141	145										
55%	111	114	117	120	124	127	131	134	138	142											
60%	111	114	117	121	124	127	131	134	138												
65%	111	115	118	121	124	128	131	135													
70%	112	115	118	121	124	128	131														
75%	112	115	118	121	125	128															
80%	112	115	118	121	125																
85%	112	115	118	122																	
90%	112	116	119																		
95%	113	116																			
100%	113																				

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.cbs.nl/information