Discussion paper

Assessing Public Sector Productivity

The case of the courts, public prosecution service and the police in The Netherlands

Brianna Gerlach

July 2021

Content

Summary

The measurement of public sector productivity has long been a controversial topic. The lack of market price information and difficulties to accurately conceptualize outputs hamper public sector productivity measurement. Adjusting for quality changes and properly aggregating different outputs constitute further challenges. Consequently, public sector productivity calculations require numerous assumptions.

This research provides a methodology to estimate public sector productivity despite the specified challenges. Specifically, it provides productivity estimates for three Dutch public sector institutions: the courts, the public prosecution service and the police. In the productivity calculations of the police, the novel concept of social costs of crime is deployed to account for differences in the magnitude of criminal activity. The results imply decreasing productivity of the courts and the public prosecution service. For the police, productivity estimates are ambiguous as the results are highly sensitive to measurement choices. While for the courts, rather reliable productivity estimates can be derived, the estimates for the public prosecution service suffer from missing information concerning the institution's cost structure. For the police, severe conceptual problems inhibit arriving at satisfactory output indicators and as such the computation of reliable productivity estimates.

Keywords

Public sector, productivity, courts, public prosecution service, police

1. Introduction

Productivity, defined as the ratio of output to input, is one of the key concepts of economics and an academically renowned indicator for economic growth. As such it is considered to be one of the most important determinants of living standards. While at first glance simply calculating the ratio of output to input may seem straightforward, estimating productivity is a much more complex endeavour. The approximation of outputs requires the use of multiple (often arbitrary) assumptions. This is valid for sectors such as market services (Inklaar, Timmer and van Ark, 2008) and especially the government or public sector. Nevertheless, public sector productivity measurement is essential as Hatry (1978:28) states: "*Unless you are keeping score, it is difficult to know whether you are winning or losing. This applies to ball games, card games, and no less to government productivity for specific services and activities*".

The challenges among practitioners and researchers to accurately measure productivity in a public sector context derive from two aspects: (1) the difficulty to conceptualize government output and (2) the lack of available price information as governments provide their services either at subsidized prices or free of charge. Therefore, the output = input convention was long the default procedure to estimate public sector productivity. This convention assumes that outputs are equal to inputs and consequently, productivity is always equal to 1 (Atkinson, 2005). However, this method has proven to be insufficient because the public sector is a large and crucial part of the Dutch economy as an employer, provider of services, consumer of resources and contributor to the economy (CPB, 2018; Atkinson, 2005; Thornhill, 2006). By assuming zero productivity growth in the public sector and not accurately "keeping score", productivity changes in a large part of the national economy are neglected. Further, actively and correctly measuring public sector productivity is crucial to evaluate public sector policies and assess public sector efficiency to implement potential improvements. As such the relevance of the topic of public sector productivity measurement is twofold: (1) economists are interested in making inferences about the effectiveness and efficiency of the public sector and (2) statisticians are interested in the correct measurement approach to derive accurate public sector productivity estimates.

When the Atkinson Review (2005) initially revived the discussion about alternative methods on how to assess productivity in the public sector and suggested using cost weights in aggregation, other scholars such as Diewert (2011; 2012; 2017) soon followed with concrete methods on how to evaluate public sector productivity more accurately assuming that data on output quantities is available. Atkinson (2005), as well as Schreyer (2012), point to the relevance of outcomes besides outputs in productivity measurement. They argue that consumers value outcomes rather than outputs. While Schreyer's (2012) analysis focuses on the health and

education sectors, insights from his study will be discussed in the adjustment of output measures for quality changes in the present study.

Going beyond a theoretical discussion of measuring productivity in the public sector, the paper at hand proposes methodologies to estimate the productivity of three Dutch public sector institutions. This paper provides productivity estimates for the Dutch judiciary, namely the courts and the public prosecution service, and the police. In contrast to previous studies, this study investigates the productivity of the courts and public prosecution service separately. Due to increasing input volumes and decreasing output volumes in almost all case categories, the productivity of the courts decreased between 2005 and 2019. This result holds no matter whether appealed cases are counted as outputs of both, first instance courts and higher appeal courts, (Total factor productivity (TFP) -15 percent) or are considered outputs of higher appeal courts only (TFP -14 percent). Due to available cost information, productivity estimates for the courts can be obtained using only a few assumptions. The final output of the public prosecution service is comprised of the cases brought to trial by prosecutors where the quality adjustment considers the number of acquittals, the cases handled by prosecutors without the intervention of a judge, investigations and special duties of the public prosecution service. As such this study provides a more holistic picture of public sector productivity than previous studies (Blank and van Heezink, 2017; Dumaij, Niaounakis and Urlings, 2014). Due to missing consumer valuation or cost information, several aggregation options for the different dimensions are proposed. Although the change in TFP differs between -11 percent and -20 percent based on the chosen premise, it can be concluded that the productivity of the public prosecution service decreased between 2013 and 2019.

The TFP estimates concerning the police are less conclusive. This analysis not only considers the crime control dimension of police work but also prevention activities, emergency responses and activities that maintain public order. While most of the presented scenarios suggest a decrease in productivity due to falling numbers of crime control output indicators which cannot be counteracted by the increase in preventative activities, the assumption of an increase in maintaining public order activities suggests an overall increase in productivity. The novel approach applied in this study of police productivity compared to previous literature (Blank and van Heezink, 2017; Urlings, 2012) is the incorporation of the concept of social costs of crime to society (Heeks et al., 2018) to account for differences in the seriousness of criminal activities. In the case of the police, the change of societal outcomes is of more interest than the change in outputs. However, the contribution of the police changes in the degree of safety is difficult to measure and quantify. This greatly inhibits arriving at reliable productivity estimates.

A common difficulty in all three analyses is to apply appropriate quality adjustments to investigate whether the decrease in output is met with an increase in quality and to incorporate the aspect of changes in case complexity which could impact productivity estimates. For the judiciary sector, explicit quality adjustments are applied, however only for certain outputs as data availability impedes quality

adjustments for all output measures. For the police, the severe conceptual problems hamper incorporating explicit quality adjustments at all. Overall, the presented case studies elucidate that productivity estimates significantly differ based on the chosen assumptions concerning output activities and their aggregation. It also points to the importance of either consumer valuation information or cost structures to aggregate outputs and stresses a potential pathway for future research in terms of estimating changes in outcomes to better evaluate the productivity of the police.

In the following, chapter 2 of this paper first depicts in greater detail the concept and measurement of productivity before elaborating on productivity measurement specifically in the government sector in chapter 3. Here, the relevance of estimating productivity in the public sector is discussed. Subsequently, this paper also reviews the caveats of the output=input convention and the challenges that arise when assessing public sector productivity. While this paper first theoretically evaluates potential methodologies on how to approximate public sector productivity when output quantities are available, it also develops a methodology to derive productivity estimates for the Dutch courts, public prosecution service and police in chapter 4. Lastly, chapter 5 concludes by providing a concrete recommendation on whether the deduced methodology is suitable for solely academia or also statistical agencies such as the Centraal Bureau voor de Statistiek (CBS).

2. The concept of productivity and its measurement

When comparing countries or firms, as well as estimating economic growth, productivity is one of the most common indicators applied among researchers and economists (OECD, 2001). Generally, productivity assesses how efficiently inputs such as labour and capital are leveraged to produce a certain output (Schreyer and Pilat, 2001). Hence, it can simply be modelled as the ratio of outputs to the required inputs as portrayed in equation (1):

$$Productivity = {}^{volume\ of\ output}/_{volume\ of\ input} \tag{1}$$

The concept of productivity measurement in economic theory has its origin in the works of Jan Tinbergen (1942) and Robert Solow (1957) who developed aggregate production functions and empirically linked these to the economic growth of countries (Jorgenson, 1995; OECD, 2001). Especially renowned in this context is Solow's (1957) growth accounting approach which decomposes output growth in the form of growth in gross domestic product (GDP) or gross domestic output into the different inputs that contribute to the growth (Schreyer and Pilat, 2001). Within

this growth model, labour and capital services function as input variables. Hence, the growth accounting equation (2) is as follows:

$$d\ln Q/_{dt} = s_L \, d\ln L/_{dt} + s_K \, d\ln K/_{dt} + d\ln A/_{dt} \quad (2)$$

The change in output *Q* is accounted for by changes in labour input *L* and changes in capital input *K* (Schreyer and Pilat, 2001). Both input changes are multiplied by their share in total costs (Schreyer and Pilat, 2001). Additionally, the so-called total-factor productivity (TFP) contributes to output growth in this productivity model. The remaining difference in output Q, which is the residual, that does not stem from differences in inputs is attributed to differences in TFP (variable *A* in equation (2); Kohli, 2010). Intuitively, this residual shows the shift in the production function (Hulten, 2000). This suggests that differences between countries in terms of income levels, living standards and economic growth prospects can be ascribed to some extent to differences in productivity (Hulten, 2009). Hsieh and Klenow (2010) report that around 50 percent (or more) of cross-country income differences cannot be attributed to differences in human and physical capital and thus have to be attributed to differences in the residual, TFP, which further underlines the significance of productivity for governments and society.

Measuring productivity can have many objectives such as possibly tracking technological change, tracking efficiency change but also measuring living standards and economic growth and comparing these across countries (OECD, 2001). In the context of firms, productivity and productivity growth are of great relevance as productivity differences are associated with firm heterogeneity. Only the most productive firms in an industry will be able to remain competitive and enter foreign markets while less productive firms will be limited to operate in the home market or even have to exit the market (Syverson, 2011; Melitz and Trefler, 2012; Helpman, Melitz and Yeaple, 2004). Nevertheless, economies and governments also try to achieve high and increasing productivity because this is one of the main approaches to reach larger economic growth and consequently increase national living standards (Lau, Lonti and Schultz, 2017).

2.1 Distinctive types of productivity measurement

Although the notion of productivity presented in equation (1) seems simple, the concept can be measured in various ways depending on the measurement of outputs and which inputs are utilized (OECD, 2001; Schreyer and Pilat, 2001; Diewert and Nakamura, 2005).

2.1.1 The output variable

The term ‘outputs’ refers to the goods and services produced (Lau, Lonti and Schultz, 2017). While in the case of services, outputs are the number of constant-quality activities or actions, in the case of goods, outputs describe the produced constant-quality physical units (Schreyer, 2012). Output quantity estimates in

productivity measurement either make use of the directly observed quantity output series or the more common method is to use current price-series or indices of output to construct the quantity measure. The quantity measure is then derived by deflation, meaning the current price series or index is divided by a price index (OECD, 2014a; Schreyer and Pilat, 2001). The quantity indices of output have to be independent of the quantity indices of inputs to ensure the validity of the resulting productivity estimate (Schreyer and Pilat, 2001). In terms of the output variable, it has to be differentiated between gross output-based approaches and value-added based output approaches (Cobbold, 2003). While the former refers to gross output that is generated by the use of all inputs, the latter additionally subtracts intermediate inputs from the generated output and thus refers to the creation of value-added (Simpson, 2009).

2.1.2 The input variable

'Inputs' refer to the factors utilized in the production of outputs. These are generally labour, capital inputs and intermediate inputs (OECD, 2001). A distinction can be made based on the type of inputs used to generate output. Most commonly this refers to either labour productivity or capital productivity. The former describes the productivity of labour for the generation of gross output or value-added output (OECD, 2001). Labour is usually one of the most critical factors for productivity and in the assessment of living standards (OECD, 2001). Therefore, evaluating labour productivity is particularly relevant to policymakers.

Measuring labour inputs can sometimes be problematic due to different measures of employment. Methods can reach from a simple headcount of employees to the most common method of quality-adjusted number of hours actually worked (Schreyer and Pilat, 2001; OECD, 2001). Appendix A supplies a more detailed discussion of labour inputs.

Capital productivity describes the productivity of capital in the generation of gross output or value-added (OECD, 2001). While capital inputs are in many aspects similar to labour inputs, there are still important differences. Capital goods are often owned by the producer and thus the 'service' provided by the capital good is not being documented as a market transaction and rents have to be imputed (Hulten, 2000). Due to the implicit nature of these transactions, measuring the quantity - The services delivered by the capital good in a certain time frame - and the pricing - The user or rental cost of capital goods - tends to be difficult (Schreyer and Pilat, 2001).

Labour productivity, as well as capital productivity, are single-factor productivity measures, as only a single input factor is linked to the output measure (Schreyer and Pilat, 2001). In contrast, in the case of total-factor productivity (TFP or MFP[1]), a range of inputs, for example, labour, capital and intermediate goods, is used to

1 Note that OECD (2001) and Schreyer and Pilat (2001) use MFP and TFP interchangeably while Diewert and Nakamura (2005) consider them as different productivity types and define MFP as productivity measure when multiple inputs are leveraged and TFP as ratio of total outputs to total inputs.

generate a certain output (OECD, 2001). TFP is a particularly helpful tool to make inferences about the contributions of labour, capital, intermediate inputs and technology to growth and hence, is important to assess past and also future economic growth (OECD, 2001). TFP growth is recorded when the growth rate of gross output increases faster than the respective growth rate of the utilized inputs (Schreyer and Pilat, 2001; Diewert, 2017; Gilder, 1975). TFP growth is mainly caused by three possible factors: (1) technical progress meaning that the production function shifts outward, (2) a combination of increasing returns to scale as well as input growth, and (3) more efficient allocation of resources from less to more productive producers (Diewert, 2017). Note that contrary to common belief TFP changes should not be equated with technological change (Diewert and Nakamura, 2005). The shift in the production function can be caused by several factors which include not only technological innovation but also changes in factor shares (Hulten, 2000). Further, it has to be distinguished between productivity levels and productivity change. A high productivity level does not automatically imply high productivity growth and vice versa (Diewert and Nakamura, 2005). This can be illustrated with the example of the so-called "Asian Tigers", South Korea, Taiwan, Singapore and Hong Kong, which experienced significantly higher growth rates than other Asian countries even though they had comparable productivity levels in the 1960s (Nelson and Pack, 1999).

2.2 Index number theory and productivity[2] measurement

Measures of outputs and inputs are critical to estimate productivity and productivity growth. However, simply summing up goods and services is impossible due to their inherent heterogeneity (McLellan, 2004). Therefore, statisticians often utilize index number formulas to aggregate several inputs and outputs respectively. Appendix B provides some background on index numbers in productivity measurement. The four most commonly used index number formulas are the Laspeyres-, Paasche-, Fisher- and Törnqvist-Theil formulas (Schreyer and Pilat, 2001). Since the choice of index formula influences the price indices and consequently the productivity estimates (Schreyer and Pilat, 2001), it is critical to choose a fitting index number formula as aggregation results may differ based on the chosen formula. Diewert (1992; 2011; 2012) advocates for using the Fisher's price index[3] as here the upward bias in the Laspeyres index formula is equalized by the downward bias in the Paasche index formula. Additionally, the Fisher price index is the only index formula out of the four mentioned above, that passes all four common axiomatic tests[4], which refer to properties that an index should show (Table C.1, Appendix C) (Diewert, 1992; McLellan, 2004). Next to the statistical approach presented above, the economic approach to index number theory should be considered. This approach accounts for the link between index number formulas and their underlying functions, for example, production functions. Index numbers

2 Using index numbers to estimate productivity was first introduced by Jorgenson and Griliches (1967).
3 The Fisher's price index is the geometric mean of the Laspeyre's and Paasche index: $P_F = [P_L * P_P]^{0.5}$.
4 For an extensive discussion of axiomatic tests and an evaluation of index formulas against different axioms refer to Diewert (1992).

that are consistent with a "flexible"[5] aggregator functional form are termed "superlative" indices (Diewert, 1976). Within the group of superlative indices, the Törnqvist-Theil price index as shown in equations (3) and (4) (Diewert, 2017) is of special interest:

$$P_T^t = \exp[\sum_{i=1}^{N} ½ \left(s_i^1 + s_i^t\right) \ln(p_i^t / p_i^1)] \quad (3)$$

$$\text{with } s_i^t = p_i^t x_i^t / p^t x^t \quad (4)$$

where p_i^t and x_i^t denote the price and quantity of inputs *i* in period t= 1,…t. The Törnqvist-Theil index is one of the most commonly used superlative indices as it is consistent with the Translog aggregator function outlined in equation (5) (Diewert, 1976) which is often used in econometrics and especially in productivity measurement[6] (Caves, Christensen and Diewert, 1982).

$$lnY = \alpha_0 + \sum_{j=1}^{N} \alpha_j \ln x_j + ½ \sum_{j=1}^{N} \sum_{k=1}^{N} \Upsilon_{jk} lnx_j \, lnx_k \quad (5)$$

Compared to the Cobb-Douglas or the Constant Elasticity of Substitution production function, the Translog production function does not rely on inflexible assumptions such as constant elasticities of substitution or constant returns to scale (Kim, 1992). In contrast, it allows for variation in elasticities of substitution among inputs when there is variation in input proportions (Dean, Harper and Sherwood, 1996). Diewert (1992) finds economic justifications for the use of both indices and thus it is possible to advocate in favour of the Fisher, as well as the Törnqvist-Theil index. It may be noted that the results when using one or the other do not differ much, which leads the OECD (2001) to conclude that there is no reason to favour one index over the other.

3. Productivity measurement in the public sector

According to the OECD (2014b), the public sector encompasses "*the general government sector plus all public corporations*". Compared to the private sector, productivity measurement in the government/public sector has to be conducted differently due to certain particularities. Nevertheless, the assessment of productivity in the public sector is critical due to a variety of reasons which will be discussed below.

5 A flexible aggregator refers to those functions which can provide a second-order approximation to an arbitrary production, cost or utility function (OECD, 2005).

6 Consult Gollop and Jorgenson (1980) for further explanation concerning the usage of translog functions in productivity measurement.

First of all, the public sector is important to the functioning of the economy overall. For example, the health and education sectors contribute to human capital deployed also in the private sector by ensuring a healthy, well-educated and consequently productive workforce (Foxton et al., 2018). Similarly, a well-functioning legal system supports the commercial sector. Hence, monitoring public sector productivity is crucial because of the connection of public and private sector where as a consequence an unproductive and inefficient public sector has potential negative spill over effects that influences the productivity of the overall economy (Foxton et al., 2018).

Secondly, the productivity of the public sector is relevant since the sector is a major consumer of tax revenues (Thornhill, 2006). The Centraal Planbureau estimated that net government expenditure was equal to 38.1% of GDP in 2018 (CPB, 2018). As such it is a topic of significant public interest whether taxes and other government income are put to good use as the productivity of the public sector as well as the efficiency of public spending significantly impacts society (Atkinson, 2005). While productivity only describes the ratio of outputs to inputs, consumers (in this case citizens), and the government, are also interested in how efficiently outputs are produced, or more specifically how efficiently tax money is spent. Citizens have increasingly higher expectations of what the public sector should deliver, but are generally unenthusiastic about paying higher taxes (Foxton et al., 2018). Consequently, the efficiency of public sector productivity is required to increase.

Increasing efficiency in the public sector indicates higher social welfare as well as better use of public expenditures (Somani, 2021). Changes in efficiency can be derived from observing changes in productivity over time. An increasing output level, given the level of inputs as well as a decreasing level of input given the level of output, implies an increasing efficiency (Mandl, Dierx and Ilzkovitz, 2008). Hence, productivity measurement is also vital to assess the efficiency of the public sector. More generally, productivity measurement finds application in evaluating public sector policies and thus, is an important tool to recognize inefficient public policies where improvements are required (Hatry, 1978).

Furthermore, Thornhill (2006) states that productivity measurement is relevant as the public sector is a major employer and a major provider of services. For many services, the public sector is the main or even sole producer, for example in the defence or infrastructure sector (Lau, Lonti and Schultz, 2017). Due to the absence of competition and the lack of automatic resource allocation from less to more productive firms present in the private sector (Melitz and Trefler, 2012) the public sector is not extrinsically forced to supply services efficiently and respond to consumer's demands (Fels and Biggar, 2017). However, the services provided by the public sector significantly contribute to the wellbeing of citizens, which is why they should be of high quality and productivity (Fels and Biggar, 2017). Consequently, it is important to actively monitor public sector productivity as market forces are missing (Dunleavy and Carrera, 2013).

Next to being a consumer of government resources, the public sector also significantly contributes to the economy as a producer (Atkinson, 2005). According to Dunleavy (2017) around 20 to 25 percent of the final economic output in OECD countries is produced by the public sector. Nevertheless, national productivity estimates do not account for productivity changes in the public sector. As these changes in productivity of a significant part of the economy (the public sector) are disregarded, this might result in incorrect productivity estimates. For example, Elbourne and Gratska (2016) report that The Netherlands have experienced rather low productivity growth after the financial crisis but including correct assessments of public sector productivity might lead to different and possibly more accurate conclusions.

To account for the output and productivity of the public sector, the Systems of National Accounts 1993 and 2008 proposed the output=input convention if direct information on output prices and quantities were not obtainable (Diewert, 2011). Here, public sector productivity is determined by stating that output (growth) is equal to input (growth), which essentially implies that productivity is equal to 1 (and growth is equal to 0) (Dunleavy, 2017).

However, the output=input convention has been criticized for several reasons. First, the convention does not incorporate any productivity changes (i.e., productivity growth) that could arise for example, due to new technologies. Therefore, government output is not properly recorded in the national accounts and as a result, the growth rate of GDP might not be correctly estimated. For example, disregarding an increase in public sector productivity would result in an understatement of GDP growth (Atkinson, 2005). Although estimating public sector productivity by using the output=input convention seems straightforward, another challenge is related to the measurement of capital services as inputs. In the System of National Accounts, capital services in government sectors are solely measured by taking into account depreciation (Diewert, 2017). Hence, the opportunity costs that arise from the foregone profit opportunities, as the capital is tied up, is disregarded (Diewert, 2017). If capital inputs are not properly measured, inputs cannot be regarded as an accurate proxy for outputs which renders the output=input convention faulty.

Due to these shortcomings, alternative methodologies have been proposed: Atkinson (2005) argues that productivity can be measured by looking at public sector outputs in terms of activities and weighing these in terms of administrative costs instead of prices which are mostly not available in the public sector. The report advocates for using direct government output measures where possible and states requirements that a directly measured output indicator needs to fulfil: *'(i) it covers adequately the full range of services for that functional area, (ii) it makes appropriate allowance for quality change, (iii) the effects of its introduction have been tested service by service, (iv) the context in which it will be published has been fully assessed, in particular the implied productivity estimate, and (v) there should be provision for regular statistical review.'* (Atkinson, 2005: 48)

Note that in this context the distinction between individual services and collective services matters. The former concern government-produced services that are supplied to and consumed by single households (OECD/Eurostat, 2012). Examples are health or education services. The latter refers to services that are simultaneously supplied to all households and which are collectively (not necessarily consciously) consumed, for example, the services provided by the police or defence sector (OECD/Eurostat, 2012). Since 2006, using output measures for individual services that are based solely on input measures (i.e. output=input convention), is not permitted anymore (The Commission of the European Communities, 2002). Nevertheless, establishing direct output measures proves to be difficult when the government output is provided in the form of collective services. Therefore, in that specific case, the output=input convention is technically still acceptable according to the Eurostat (2016) Handbook on prices and volume measures in national accounts.

Although the output = input convention is generally deemed an insufficient productivity measure, this feeling is not universally shared. Lynch (2006) argues in its favour and criticizes the principles presented in the Atkinson Review (2005). He states that the provision of services by the public sector for citizens cannot be compared to a market transaction as there is no economic exchange. He proclaims that the government consumes inputs in the name of society and then provides the produced services to the entire society. As such the government does not provide an output that is consumed by society and should not be distinguished in national accounts[7] (Lynch, 2006).

3.1 The challenges in measuring productivity in the public sector

Estimating productivity in the public sector bears multiple difficulties. The poor availability of information on prices for government outputs, the conceptualization of outputs and quality adjustments pose a particular problem.

3.1.1 The challenge of conceptualizing output

An essential step to calculate productivity is to obtain an output series. However, this is where the first challenge emerges. How does one conceptualize output and consequently measure it accurately? What exactly constitutes a unit of output for public sector institutions? This is a complex question to answer in the context of the public sector[8]. In particular, arriving at observable output series is difficult for collective services such as policing or defences where it is not always obvious how

7 Lynch (2006) compares this to the situation of a housewife who buys on behalf of her household, then provides the inputs and her labour to the household but this is not portrayed as output provided to the household.

8 It should be noted that this problem is not exclusive to the public sector but the conceptualization of outputs can also be problematic in the private sector, for example in the context of the financial sector (Inklaar, Timmer and van Ark, 2008).

to conceptualize outputs (Caplan, 2006; Atkinson, 2005). In the case of sectors providing individual services, actual outputs that are consumed by consumers exist, for example, the amount of teaching consumed by a student (Atkinson, 2005). Consequently, it might be possible to use measurable activities such as the number of lessons taught in a school to obtain volume output measures (Simpson, 2009). For collective services, conceptualizing output is less straightforward due to their nature. As these services are provided to many consumers simultaneously a direct transaction between provider and consumer is missing. The New Zealand Productivity Commission (2018) suggests using indicators of activities such as the number of hours spent on fire prevention as a proxy. Observing activities provides information about what public sector institutions actually use their inputs for and can thus be regarded as being a close approximation to outputs (Atkinson, 2005). It is important to highlight that a count of activities does not always reflect the entirety of the outputs of a public sector unit, especially if it is difficult to pin down what exactly the output of a public sector unit encompasses(Simpson, 2009).

In the context of this discussion, it is necessary to point to outcomes besides outputs. An outcome is a situation that consumers value – for example, a certain level of knowledge or a certain level of safety (Schreyer, 2012). The difference between output and outcome is that outputs are the produced goods and services while outcomes refer to the purpose where goods and services are applied. Other than the considered outputs, outcomes might be impacted by other factors[9] (Atkinson, 2005). Although the two concepts are distinctive, they cannot be regarded completely independently since according to Atkinson (2005) government outputs should be adjusted for quality to derive accurate productivity estimates. This reflects that the neglect of accounting for quality changes may result in misleading effects such as over- or understatements of productivity[10] (Atkinson, 2005). For example, suppose hospital A discharges a patient after two days but has to readmit said patient again after a day and then discharges him again. Technically, hospital A produces more output than hospital B that discharges a fully healed patient after four days (Dunleavy, 2015). Nevertheless, in terms of the quality of the output, hospital A produces lower-quality output than hospital B. Similarly, a decrease in output not necessarily indicates a decrease in productivity but can also be caused by an increase in the quality of outputs and as such, point to a productivity increase (Zouridis et al., 2014).

In the private sector, price increases can be regarded as an indicator of increasing quality. Additionally, if a consumer is willing to pay a higher price for a product this also indicates higher welfare derived from the product (Dunleavy, 2015; 2017). However, in the public sector price information is often not available and additionally, quality changes are not necessarily reflected in price changes (Diewert, 2017). Although cost information is available in the public sector and higher-cost services might seem to be of higher quality than lower-cost services, equating costs

9 For example, the health status of a population may also be influenced by dietary and exercise habits or the incidence of smokers in the population and not only the work of the health sector (Atkinson, 2005).

10 Gemmel, Nolan and Scobie (2018) show with their study of productivity in the tertiary education sector in New Zealand that adjusting for quality can result in different conclusions concerning productivity trends.

with quality is also not ideal (ONS, 2019). For example, if due to technical innovations a more cost-effective new laser treatment becomes available which requires less recovery time than the old treatment, this can be considered an increase in quality but no corresponding price- or cost increase is evident (Schreyer, 2010). Due to the missing prices, it is further not possible to observe possible welfare changes arising from the shorter recovery time of the new treatment. As methods similar to the ones used in the private market, cannot be deployed in adjusting public sector outputs for quality, this is where outcomes come into play.

Schreyer (2010; 2012), Atkinson (2005), as well as the ONS Guide to quality adjustment in public service productivity measures (2018), point to the importance of outcome measures in the context of adjusting for quality change by considering to what extent the service outputs contribute to outcomes. In the example of the two hospitals above, the quality can be judged by the outcome (i.e. discharging a healthy patient). Hospital A does not contribute to that outcome when it releases the patient the first time. It is important that the focus solely lies on the "*incremental impact on outcomes arising from the activities of the public sector*" (Atkinson, 2005: 41) as outcomes are also influenced by factors other than the output activities. This results in an outcome-based output approach put forward by the Atkinson Review (2005) which bases the output measure on the marginal contribution of the service to an outcome[11]. In other words: it is assumed that consumers only value outcomes and not outputs. This aspect is relevant and further discussed in the context of the police.

The view is contested by Caplan (2006). He states that including outcomes in the discussion surrounding quality adjustment is not ideal because (1) outcome measures might lack reliability, (2) there is a possible time lag between the production of the output and the caused changes in outcome (3) changes in outcomes can also be caused by other factors besides the output and (4) outcome measures might not be able to capture all aspects of quality in public services.

3.1.2 The challenge of missing price information

The second challenge is the lack of pricing in the public sector. Public sector services (and goods) are often either free of charge or supplied at much lower and subsidized prices, meaning that market prices are not available or not representative (Diewert, 2017; Lau, Lonti, Schultz, 2017). Since productivity measurement using index number theory requires output quantities as well as prices, the absence of prices for government outputs poses a major problem (Simpson, 2009). First, the valuation of goods and services proves problematic without price information. In a perfectly competitive market, prices serve as a proxy for relative marginal valuations of the provided goods and services (Simpson, 2009). However, in the public sector, these market transactions are missing and hence, it is more difficult to properly value the provided services (Atkinson, 2005). The lack of market prices thus impedes the aggregation of outputs to arrive at a total

11 An example for this approach is the work by Jorgenson and Fraumeni (1989) who measure output in the education sector by investigating the increase in human capital due to education.

volume valuation. In the private sector, the relative valuations of outputs based on market prices can function as weights and thus account for differences in valuations of outputs in aggregation (Simpson, 2009; Dunleavy and Carrera, 2013). In the public sector, prices acting as the basis for these weights are missing and thus aggregating multiple output series is difficult. Consider the following example by Annabel (2019): while it may seem that a brain surgery conducted in a public hospital provides higher value than a hernia repair, there are no prices attached to the procedures and therefore, the relative value of the brain surgery to the hernia repair cannot be estimated. If then the two outputs are supposed to be aggregated to measure the total output of the hospital, it is not clear how much more value to attribute to the brain surgery compared to the hernia repair in aggregation. Since prices, which function as a proxy for marginal consumer valuation in the private sector are not available, the most common output aggregation method is to use average unit costs also called cost weights (Lau, Lonti and Schultz, 2017; Atkinson, 2005; Diewert, 2012). This practice shows the supply side instead of the valuation by society. Hence, Atkinson (2005) criticizes this practice and advocates in favour of using marginal valuations as weights in aggregation. If firms in the market sector provide services comparable to the public sector institution, their prices may be deployed as weights. However, often only costs are available and using cost-weights is the only feasible option (Atkinson, 2005).

3.2 New approaches to measuring productivity in public sectors

3.2.1 Measuring productivity when output quantities are available

Although missing price and quantity information impede public sector productivity measurement, Diewert (2011; 2017) proposes several methods that go beyond solely applying the output=input convention. Nevertheless, these methods require a minimum amount of information on nonmarket output, namely on output quantities. The best-fitting method to estimate public sector productivity when output quantities are available depends on the specific type of information available. One of the most common methods proposed by Diewert (2011; 2017) is the valuation based on the cost of production[12]. Valuations based on cost functions are applicable if information on marginal costs, incremental costs or cost functions for inputs for the examined periods are available. These cost information are employed to arrive at valuations of government output (Diewert, 2011). A specific variant of the cost-based method is discussed in Diewert (2017). If output quantities are available (as well as input quantities and prices), it is noteworthy that outputs in period t have to be normalized which is required so that output in period

12 Another method uses purchaser's valuations where the weights represent the relative value of the outputs to users or recipients. The vector of weights is utilized to aggregate the public sector outputs and can also be regarded as quality adjustment factors as higher weights indicate that users value this output more than outputs with a lower weight (Diewert, 2017). The difficulty with this method is how to determine the weights as purchaser's valuation is mostly not known. A third approach uses indirect purchaser's valuations (Diewert, 2011). These methods make use of either Fisher's price and growth indices if government outputs are sold at detectable market prices or comparable prices of private sector outputs that are relatively similar to the public sector outputs (Diewert, 2011).

1 is equal to input costs in period 1 and input indices can be deployed[13] (Diewert, 2017). From this follows that the normalized output quantity Q^t for t = 1,…,t is (Diewert, 2017):

$$Q^t = \sum_{i=1}^{N} [p_i^1 * x_i^1] [{}^{q^t}/_{q^1}] \tag{6}$$

In this case, p^t denotes the price vectors for inputs in period t, x^t denotes the quantity vector for inputs and q^t denotes the output quantity produced. The implicit Törnqvist input quantity index[14] Q_T^t for t = 1,…, t can be calculated as presented in equation (7) by dividing the total input cost of each period by input price index P_T^t calculated previously in equation (3)(Diewert, 2017).

$$Q_T^t = {\sum_{i=1}^{N} p_i^t \, x_i^t}/{\mathrm{P}_T^t} \tag{7}$$

This provides all necessary components to estimate the productivity of a public sector unit (despite missing output prices) for t=1,..,t as follows (Diewert, 2017):

$$TFP_T^t = {Q^t}/{Q_T^t} \tag{8}$$

If information on the fraction of the total costs of period t that is caused by the production of output i (f_i^t >0) is available besides the information on quantity produced by the public sector unit, q_i^t for i=1,…,I and t=1,..,t the cost-based prices for several outputs in period t can be computed, denoting P_i^t with the following formula (Diewert, 2017):

$$P_i^t = f_i^t \, (p^t * x^t)/q_i^t \tag{9}$$

This approach provides the required quantity vectors q^t = [q_1^t,…,q_i^t] and price vectors P^t = [P1t,…,Pit] to arrive at output aggregates[15]. Since input, as well as output quantities and prices, are available, normal index number theory can be applied. However, the drawback of this approach is that the total costs of production have to be decomposed into the fractions caused by each output (Diewert, 2017).

3.2.2 Incorporating quality changes

As discussed above, to properly compare productivity over time, output measures have to be quality-adjusted. Although several advances have been made towards the incorporation of quality-adjusted output measures, especially for sectors such as health and education[16], this aspect still poses challenges for many other sectors.

---

13 The normalization of output is further necessary when calculating an independent period t cost-based output Price P^t for each period t (Diewert, 2017): $P^t = {\sum_{i=1}^{N} [p^t * x^t]}/{Q_t}$.

14 Similar calculations can be conducted with the other 3 indices.

15 Following Diewert (2017) these output aggregates also have to be normalized as shown in equation (6).

16 See Clemmensen (2018), Gemmel; Nolan and Scobie (2017), Caplan (2006), ONS (2019).

Therefore, the following section will introduce several quality-adjustment approaches but the reader is asked to keep in mind that this issue still requires further research (Schreyer, 2012; Atkinson, 2005).

One possibility to implicitly adjust for quality change is called “stratification” (Schreyer, 2010; 2012) or “differentiation” (Atkinson, 2005). Stratification entails that goods and services are categorized or matched together based on their quality meaning that services of different quality are regarded as being part of different categories. Therefore, a comparison over time in terms of quantity and price changes is only drawn for goods and services of that specific category. Hence, it only compares products of similar quality and thus observes price changes while the quality is kept constant (Schreyer, 2010; 2012). If the obtained categories are homogeneous, the quality change is evident through a shift in the proportion of different services in total services. This means that if consumers opt for higher quality services the share of the respective service in total service will increase and a change is visible (Atkinson, 2005). Nevertheless, this approach can only partly adjust for quality and it disregards substitution processes, for example, due to improved technologies (Schreyer, 2010). Going back to the previous example of the old and new (more cost-effective) treatment: If the two treatments are aggregated into a single output indicator according to cost weights which is proposed by Atkinson (2005), an increase in new treatments causes a decrease of the aggregated output indicator (Schreyer, 2010). Nevertheless, consumer valuation has to be considered in this scenario. In the case that consumers regard both treatments as perfect substitutes and do not prefer one to the other, cost weights do not need to be applied and the volume index is derived by adding up the number of treatments of both kinds (Schreyer, 2010). If consumers are not indifferent between the two treatments, an explicit quality adjustment has to be applied. This can be achieved either by upwards re-scaling of the new treatment quantity or downwards re-scaling of the old treatment quantity, but whichever way is chosen, one treatment is expressed in equivalents of the other (Schreyer, 2010). This ratio is supposed to indicate consumer preferences towards the treatments, however, this is also what makes this approach difficult to implement as information about consumer preferences has to be available (Schreyer, 2010). Consumer preferences differ according to different outcomes of the two treatments (i.e. the new treatment requires less recovery time).

This ties in with the discussion of section 3.1.1 that outcomes should to some extent be considered in the quality-adjustment of output measures. A more direct approach to quality adjustment building on outcomes is if changes in outcomes are related to changes in outputs. By measuring the changes in outcome, explicit quality adjustment as presented above is not required anymore (Atkinson, 2005; Schreyer, 2012). Cutler et al. (2021) provide a recent example of this approach. The authors measure the changes in medical spending and relate this to changes in health outcomes for 80 medical conditions (e.g. heart diseases, cancer and frailty). The novelty of Cutler et al. (2021) is that they differentiate between the medical conditions which are used as the “industry” rather than the providers such as hospitals or pharmacies. Consequently, they account for possible substitutions

between medical providers and their study focuses on medical spending which is the input and the resulting changes in outcomes in terms of health.

Furthermore, it is possible to account for quality changes in an explicit manner by using hedonic regressions[17]. This method entails that the price of a product is regressed on a vector *z* of quality determining characteristics that determine the price of said product (Diewert, 2011; Schreyer, 2010; 2012). The characteristics included in vector *z* should be of importance to consumers[18] (Schreyer, 2012). The regression coefficients obtained from a hedonic regression provide a value for the characteristics and are thus called 'implicit prices' or 'characteristics prices' (Triplett, 2006). They basically indicate the price for an increment of one unit of the characteristics in a market environment as these are determined by supply and demand (Triplett, 2006). However, in a non-market environment, where supply is not directly influenced by demand, the regression coefficients reflect costs. Therefore, the coefficients can be used to approximate the quality-adjusted rate of change in unit costs (Schreyer, 2012). Although this valuation is based on costs, which is not ideal according to ONS (2019), the characteristics included in vector *z* ideally reflect consumer valuation. However, Schreyer (2012) acknowledges that this method requires large amounts of data and has not been tested yet. Additionally, it is possible to use hedonic regressions including private sector products to determine the value of a new public sector output if the new public sector output has characteristics that are to some extent similar to the characteristics of outputs sold in the private sector (Diewert, 2017).

4. Measuring productivity for the Dutch courts, public prosecution service and police

The previous chapters shed light on the importance but also the challenges that arise when measuring productivity in the public sector. While the topic of public sector productivity has extensively been investigated theoretically, this chapter will take a more practical stance and a methodology will be presented to estimate the productivity of three specific Dutch public sector institutions, namely de Rechtspraak ('the courts'), het Openbaar Ministerie (OM) ('the public prosecution service') and de Politie ('the police'). Note that the productivity developments

17 For an extensive discussion of hedonic price indices see Triplett (2006).

18 An example of quality determining characteristics concerning the output of the justice system could be the number of appeals of decisions taken in the court, the handling time of court cases but also the number of complaints received by clients (Blank and van Heezink, 2017). Note that in this case the inverse would have to be taken as generally, a higher quantity of the quality determining characteristics implies a better quality which is the opposite for the examples given.

presented on the following pages describe changes in productivity and not levels of productivity. Although two of the three institutions show a clear negative productivity trend, in a cross-country comparison, the productivity of Dutch institutions could still be above average.

The analyses presented in the paper at hand are not the first attempts to produce productivity estimates for these three Dutch institutions. Previous analyses investigating productivity in the judiciary sector include among others Blank and van Heezink (2017), Dumaij, Urlings and Niaounakis (2014) or Pommer and Eggink (2010). Except for Pommer and Eggink (2010), these studies observe the joint productivity of courts and public prosecution service. The research conducted by Dumaij, Niaounakis and Urlings (2014) only includes cases brought before the district courts to approximate the output of the public prosecution service and thus does not consider all case types. Further, while the authors distinguish between cases handled at different instances, no distinction is made between civil, criminal and administrative cases. These are all aggregated into a single output variable. Their study shows that productivity significantly decreased between 1983 and 2002 by on average 5.3 percent annually. Between 2002 and 2011, productivity still decreased but only at 1.2 percent on average annually. Blank and van Heezink (2017) consider different court case categories, for example, criminal, civil and administrative cases, and also account for relative weighing differences between those in aggregation. In terms of the OM, they consider the number of transactions, sepots, penalties and administrative appeals as outputs. Similarly, to Dumaij, Niaounakis and Urlings (2014), Blank and van Heezink (2017) find a negative productivity trend for the judiciary, more specifically a decrease of 40 percent in productivity between 1980 and 2001.

In contrast to the other two studies, Pommer and Eggink (2010) investigate the productivity of the OM and the courts separately and find that productivity decreased annually by 1.2 percent and 2.5 percent respectively between 2000 and 2008. The authors attribute this decrease to increases in material and labour costs. The originality of the present study compared to previous studies lies in the treatment of the two institutions as separate entities and consequently providing individual productivity estimates. Further, the present study distinguishes between more case categories in the case of the courts and uses the price information available to derive cost-weights that account for changes in relative costs over time. In contrast to the treatment of the public prosecution service in the other studies, this research investigates a broader set of outputs that next to cases brought before the court and settlements also includes other dimensions of the institution's work. Lastly, it considers a more recent time period.

In the context of the Dutch police, Urlings (2012) estimates that the productivity of the police increased between 1955 and 1984 but decreased from 1985 to 2006 and showed an even stronger decrease from 2006 to 2011. The author distinguishes between crimes and traffic offences to account for differences in seriousness. Urlings, Blank and Nianouakis (2014) broaden their analysis compared to Urlings (2012) by also incorporating other dimensions of police work such as prevention.

The observed productivity trend between 2000 and 2012 is again negative. Blank and van Heezink (2017) find that after a significant increase in productivity between 1980 and 1986, a negative productivity trend is apparent afterwards. In 2011, productivity is at the level of 1980. The novelty of the approach applied in the present analysis lies in several aspects. First, this approach does not only incorporate the total number of crimes but accounts for eight different types of criminal activity. Second, these different types of criminal activity are aggregated using different weights which are derived using the concept of the social costs of crime (Heeks et al. 2018). The social costs of crime provide information about the costs a crime of certain type causes. Hereby, not only medical costs but also other costs resulting from crime are considered. This allows accounting for differences in the seriousness of crimes on a deeper level. The present study further covers a more recent period than other studies that begins after the restructuring of the police. Lastly, similar to Blank and van Heezink (2017) the present analysis aims to consider several aspects of police work to capture large parts of the production process of the police.

Inputs and Deflators

To make inferences about the inputs used in the output production process, the framework of the Classification of the functions of Government (COFOG) is a good starting point. COFOG is an international standard classification developed by the OECD which categorises data concerning public expenditure from the System of National Accounts according to the main purpose of its use (OECD, 2019). Thus, it provides the necessary inputs data for the productivity calculations. The COFOG framework includes two levels: The first level disaggregates the expenditure data into ten functions of the government such as 'public order and safety' or 'health'. These functions are further divided into several sub-categories in the second level[19] (OECD, 2019). In the COFOG framework, total government expenditures or inputs for each sector are composed of the following categories: intermediate consumption (+), employee remuneration (+), interest in accordance with national accounts (+)[20], social benefits (+)[21], subsidies (-), other current expenditures (+), other capital expenses (+) (Eurostat, 2019). Nevertheless, the data from COFOG is not suitable for all public sector institutions. In some instances, the data is not sufficiently disaggregated. In that case, the available inputs data will be adjusted to fit with the COFOG framework as much as possible.

The COFOG database supplies nominal costs for the input categories listed above. Nevertheless, the estimation of productivity requires input quantity indices instead. In some instances, volume measures for inputs are readily available. This is often the case for labour inputs where (if available) full-time equivalents (FTEs) provide a direct measure of input volumes. The other method to obtain input quantities requires the translation of nominal input costs into volume measures. This can be done by using price indices to deflate the cost figures (OECD, 2014a). Here, suitable

19 For example, 'Public order and safety' is split into the functions of (among others) 'police' and 'prisons'.
20 There is no data for this category for police, OM or the courts as it is part of labour compensation.
21 Again there is no data for this category for the examined institutions as this is included in labour compensations.

deflators to be used in the analysis need to be chosen[22]. The deflators used in this analysis are the following: the Index Materiële Overheidsconsumptie ('price index for the material consumption of the government') for intermediate consumption (CBS, 2021e), the Index Bruto Overheidsinvestering ('price index for gross government investments') for capital expenses (CBS, 2020a) and the Consumer Price Index for taxes and other current expenses (CBS, 2021b). Labour inputs that are not given in direct labour measures will be deflated using a deflator calculated from the input-output tables as well as sector accounts. Appendix D provides an elaboration on the deflators and a more detailed explanation of the labour deflator.

4.1 The courts

4.1.1 Overview

The Dutch system of government follows the Trias Politica system proposed by Montesquieu in 1748 which postulates that there are three branches of government (Pot, 2016; Montesquieu, Nugent and Alembert, 1899). The third of these branches is concerned with the settlement of conflicts. This position is filled by the courts. Together with the public prosecution service, the courts form the Dutch judiciary whose purpose is to provide "*an independent delivery of Justice*"(Council for the Judiciary, 2010: 5). The Courts are responsible for truth-finding and conducting trials to settle conflicts. They are under the authority of the Ministry of Justice and Safety but hold an independent position (Pot, 2016). In The Netherlands, several different categories of courts exist. These are organized in several instances. Eleven Rechtbanken ('district courts') handle cases concerning civil law, criminal law, cantonal matters and administrative law in the first instance (De Rechtspraak, n.d.b.). If a client does not agree with the verdict handed down at one of these courts, the client can appeal to a court of higher instance. The case is then handled at one of four Gerechtshoven ('Courts of appeal'). The courts of appeal deal with appeals in criminal, civil, cantonal and tax matters. The Hoge Raad ('supreme court') is the highest judicial instance in The Netherlands (De Rechtspraak, n.d.b.). It has the power to determine that a case is closed or has to be reopened. Besides, the College van Beroep van het Bedrijfsleven (Cbb) ('Board of Appeals of Business') handles cases concerned with business matters such as the Competition act. The Raad van State ('Council of State') is the highest instance for administrative cases. The Centrale Raad van Beroep ('Central Board of Appeals') is responsible for higher appeal cases concerning matters such as study financing, social security law and civil servants law (De Rechstpraak, n.d.b.).

4.1.2 Inputs

As stated in the introduction to this chapter, the analysis of inputs will be based on the classification of public expenditure according to the COFOG framework. However, in the COFOG classification, the courts are part of the first-level category of Public Order and Safety and the second-level category De Rechtspraak. This

22 Atkinson (2005) points out that deflators are especially important where direct indicators for output are not available as then deflated input volumes are part of constant price GDP.

category includes the courts as well as the public prosecution service. Therefore, the available second-level data in the COFOG dataset is not sufficiently disaggregated and concerns both institutions. Theoretically, a third level providing more granular data would be required. Nevertheless, the annual reports provide information concerning the utilized inputs. These inputs have to be categorized to fit with the COFOG framework. Appendix A supplies a more detailed discussion on the aggregation of inputs, some comments concerning the direct volume measure for labour inputs and Figure A.1 displays the evolution of the court's inputs. Table 4.1.2.1 portrays the inputs used in the following analysis.

Table 4.1.2.1. Inputs, measurements, % change in (deflated) inputs and sources for the analysis of the courts

Inputs[23]	Measurement	Change 2005-2019	Source
Intermediate Consumption	Million euro	-24%	De Rechtspraak (2006-2019)
Employee remuneration	Million euro	+63%	De Rechtspraak (2006-2019)
Other current expenditures	Million euro	-52%	De Rechtspraak (2006-2019)
Other capital expenditure	Million euro	-12%	De Rechtspraak (2006-2019)
Employees	FTE	+13%	De Rechtspraak (2006-2019)

4.1.3 Outputs

Following from the courts' tasks, a reasonable proxy for the output is the number of conducted trials. These can be differentiated into ten different case categories (De Rechtspraak, 2020). The district courts distinguish between six different categories, namely criminal cases, civil cases, administrative cases, tax cases, immigration cases and cantonal cases. The courts of higher appeal deal with appeals in criminal, civil and tax cases and the last category concerns cases handled by the Central Board of Appeals. While StatLine provided detailed data concerning the production and activities of the courts, since 2016, the database has been discontinued. To cover a longer period, this analysis utilizes data from the annual reports (De Rechtspraak, 2006-2020). The number of cases handled per case category and for different instances is available in the reports from 2005 onwards. Consequently, this analysis will examine the evolution of productivity between 2005 and 2019. Table 4.1.3.1 provides an overview of the output indicators employed in the following analysis is.

Table 4.1.3.1. Output indicators, values and sources for the analysis of the courts

Output indicator	2005	2019	Source
Aggregated cases – all cases (AC_t)	1 740 460	1 510 920	De Rechtspraak (2006-2020)
Aggregated cases - excluding appealed first instance cases (FI_t)	1 679 780	1 460 270	De Rechtspraak (2006-2020)

23 The costs items available in the annual reports include labour costs, material costs classified as intermediate consumption, depreciation costs classified as other capital expenditure and so-called court costs classified as other current expenses. For 2005-2016, interest expenses are recorded but as these are not available for the entire period and only make up less than 1% of total input costs, they are disregarded.

Price discussion

An interesting aspect of the discussion concerning the outputs of courts is that "prices" for the different case types are available. Since January 1st 2005, the courts are compensated according to output financing (De Rechtspraak, 2006) and the compensation received per handled case of each category is determined by discussions between the Minister of Justice and Safety as well as the Council for the Judiciary (Pot, 2016). Although the Council for the Judiciary takes part in the price negotiations, the courts can be considered to be price takers to some extent. The price per case category is based on the historic realised prices, handling time, quality standards, workload effects (e.g. due to new legislation) and seriousness (Blank and van Heezink, 2017; Pot, 2016). This indicates that case categories where cases require (on average) longer handling times, more resources and are of higher seriousness are supposed to be allocated a higher price. However, this also implies that the compensation received by the courts is associated with the costs of handling different cases rather than consumers' (citizens') marginal valuation of different case categories. The prices are reassessed every three years (De Rechtspraak, 2020). Prices are fixed for these three years and are not adjusted during this period even if it becomes evident that actual costs differ from projections (Council for the Judiciary, n.d.).

The annual reports provide data on the estimated number of handled cases and the corresponding monetary compensation. It is noteworthy that the number of handled cases (meaning the production) is agreed-upon between the courts and the Minister of Justice and Safety before the start of the year (De Rechtspraak, 2006). At the end of the year, these estimates are compared with the actual number of handled cases. If the courts handle more cases than estimated, the payment for the excess cases will be 70 percent of the negotiated compensation for the respective case type. In the case, they handle fewer cases than expected, the courts will receive 100 percent of the agreed-upon compensation for the handled cases and 70 percent for the nominally missing cases (De Rechtspraak, 2005). Due to this payment rule for additional or reduced production, the realised prices per case category differs from the agreed-upon price. The realised prices are not published in all annual reports. Nevertheless, with the information available concerning the estimated and realised volume of cases handled as well as the prices paid it is possible to calculate the realised revenue generated by each case type (equation (10)).

$$RR_i = ER_i + (RV_i - EV_i) * 0.7 * P_i \tag{10}$$

Here, RR_i stands for the realised revenue, ER_i is the estimated revenue, RV_i is the realised volume, EV_i is the estimated volume, P_i is the price paid for each case and the subscript i denotes the different case categories. The realised price per case category paid to the different courts can easily be estimated by dividing the realised revenue (RR_i) by the realised volume (RV_i) of cases per case type. To be consistent, the realised price is calculated in the same manner for each year, including years that a realised price is published because small differences exist between the published and calculated realised price. Table 4.1.3.2 shows that the changes in

realised prices from 2005 to 2019 differ greatly between the different case categories. While the price for higher appeal tax cases increased by 304 percent, compensation for cantonal cases only increased by 1 percent.

Table 4.1.3.2. Prices and price changes per case category at the courts

Case category	2005	2019	Price change
	Euro		%
Criminal cases	606.69	1 168.41	93
Civil cases	747.91	840.96	13
Administrative cases	1 638.85	2 399.55	46
Immigration cases	976.14	1 405.37	44
Cantonal cases	116.20	165.48	42
Tax cases	1 361.99	1 375.40	1
Civil law higher appeal	3 487.99	4 359.56	25
Criminal law higher appeal	1 201.20	1 723.22	43
Tax cases higher appeal	1 001.39	4 046.54	304
Central Board of appeals	2 750.07	4 088.13	49

Two aspects require consideration. First, the aspect of quality. It may seem tempting to derive changes in quality from changes in the negotiated prices. However, although the price is based on indicators such as handling times, the annual reports do not explicitly state that price increases are associated with quality increases or to what extent changes in the negotiated prices can be attributed to apparent or expected quality changes. It could also be the case that an increase in quality is expressed in higher efficiency of the courts which could reduce handling times. In that case, price information would not reflect such increases in quality. The price changes rather reflect cost increases than changes in quality. Therefore, this analysis will not apply price information to make inferences about overall quality changes. Second, the aspect of case complexity. While most case categories display a significant price increase which can be attributed to inflation but also increasing costs of the respective case categories, changes in case complexity within a case category cannot directly be deduced from price changes. Higher-complexity cases might result in longer handing times and require more resources. These are both determinants of the negotiated prices which could imply that increasing case complexity is compensated through paying higher prices. Nevertheless, again it is not possible to explicitly attribute price changes to increasing case complexity and thus no inferences about changing case complexity will be made. In the future, this might be possible as the report for 2019 (De Rechtspraak, 2020) states that the price catalogue for the period 2020-2022 includes price changes to compensate for increasing case complexity.

4.1.4 Adjusting for quality change

As mentioned above, quality changes will not be deduced from price changes. Nevertheless, Atkinson (2005) stresses that output measures should be adjusted for quality (if possible). Measuring or finding a proxy for quality generally poses a difficulty. It is not straightforward to adjust for quality in the case of the courts.

Implicit quality adjustment in the sense of stratification as proposed by Schreyer (2012) is already incorporated in the analysis, as cases are grouped into case categories and compared over time. Nevertheless, it would be desirable to also incorporate explicit quality adjustment. The courts lay out certain quality measures for themselves. These are supposed to ensure high-quality work and decrease the focus on the number of handled cases which is imposed by output financing (Council for the Judiciary, n.d.). Examples thereof are: judges and support staff receive 30 hours per year or at least 90 hours every 3 years to further educate themselves (Netherlands Council for the Judiciary, 2008). Furthermore, the judiciary determines a minimum percentage of cases that have to be handled by a full bench (meaning three judges instead of just one). The intuition behind this measure is that discussions between several judges stimulate improvements in the legal quality of verdicts (De Rechtspraak, n.d.a). For these proposed indicators, it is difficult to find suitable data. Therefore, handling times are often suggested as a quality indicator. Although such data is available, this indicator will not be applied in this analysis because handling times are also dependent on case complexity. Thus, they do not necessarily represent changes in quality adequately. A rather novel approach would be to incorporate society's trust in judges as a quality indicator but trust data is only available from 2012 onwards. Additionally, Bouckaert and van de Walle (2003) argue that the level of trust cannot necessarily be ascribed to the functioning of an institution, since for example, events greatly influence citizens' trust in institutions although the quality of the institution remains unchanged.

In the presented analysis, the number of appeals will be used to make inferences about quality. Here, the premise is that one of the parties of the case only appeals if there is a reason to assume the appeal will be granted. Therefore, it is assumed that parties only appeal if they think a mistake has been made at the first instance court. This proposition is based on the approach proposed by the European Commission for the Efficiency of Justice (2016). The present study suggests that outputs will be calculated in two ways. The assumption underlying the first approach is that all cases count as output. This means that appealed cases are counted twice, once as output for the respective first instance courts and the second time as outputs for the appeal courts. The output indicator, in this case, is $AC_{,t}$ denoting the aggregated index of all cases in year t (equation (11)). This approach implies that more appeals constitute a larger output of the courts which, under the premise that appeals signal lower quality work of judges, would result in misleading productivity conclusions.

The second approach assumes that only the final outputs count. Hence, cases that are appealed are not regarded as the output of the first instance courts as it is assumed that a mistake was made. Appealed cases can then be regarded as intermediate products. Blank and van Heezink (2017) compare this to the situation when a customer returns a purchased product and orders a new one because the product does not satisfy expectations. The delivered product is only counted once, not twice (Blank and van Heezink, 2017). Therefore, the cases of appeals are subtracted from the number of cases at the respective court of first instance. The

respective output indicator is denoted as FI_t where t denotes the year (equation (12)).

To derive FI_t, the number of appeals handled by the criminal court of appeals is subtracted from the cases handled at the first instance criminal court to avoid the double-counting of cases. Similarly, the appeals at tax courts of appeal and civil courts of appeal are subtracted from the number of cases handled at the respective courts. The cases of the Centrale Board of Appeals are appeals of administrative cases, consequently, these are subtracted from administrative cases (De Rechtspraak, n.d.c.). While immigration cases are also generally part of the administrative category, they are regarded as an individual category in the first approach. Appeals in immigration cases are handled by the Raad van State where case numbers are not given in the annual reports and thus for that category, no double-counting of appeals is apparent and hence, no subtracting of appeals necessary (De Rechtspraak, n.d.c.).

It should be considered that information concerning the number of appeals is "to be handled with care" (European Commission for the Efficiency of Justice, 2016) since it can be the case that verdicts are appealed to cause delays of a process. Ideally, the measure of the held appeal rate should be considered. The number of held appeals provides information about the number of appeals that confirm the decision taken in the first instance (European Commission for the Efficiency of Justice, 2016). Although such data is not available for the present analysis, it would benefit the reliability of the presented analysis as the number of held appeals would more accurately reflect the quality of verdicts.

4.1.5 TFP analysis

To compute TFP, the different case categories have to be aggregated into a single output measure. As described above, Atkinson (2005) advocates for aggregating different outputs according to marginal valuations. However, only cost information is available. Under the assumption that the realised prices accurately reflect the handling times, workload and seriousness of each case category, the prices can be regarded as presenting the relative marginal costs of the different case categories. Consequently, the weights used in aggregation can be based on the realised prices. In equation (11) the volume measures of handled cases of the different categories (C^i_t) are aggregated according to their previous years' cost share (w^i_{t-1}) in total costs. Equation (12) follows a similar principle but the appeals are subtracted. The weighting slightly changes from year to year, however, it should be noted that the largest change in cost shares is at most 5 percent. Table 4.1.5.1 shows the output changes as well as the average weights per case category.

$$AC_t = \sum_{i=1}^{N} (C_t^i * w_{t-1}^i) \tag{11}$$

$$FI_t = (CrC_t - CrA_t) * wCrC_{t-1} + (CL_t - CLA_t) * wCL_{t-1} + (AD_t - BA_t) * wAD_{t-1} + IM_t * wIM_{t-1} + Ca_t * wCa_{t-1} + (TC_t - TCA_t) * wTC_{t-1} + CrA_t * wCrA_{t-1} + CLA_t * wCLA_{t-1} + TCA_t * wTCA_{t-1} + BA_t * wBA_{t-1} \quad (12)$$

Table 4.1.5.1. Average weights and output changes per case category at the courts

Case category (C_t^i)	Average weights (AC_t)	Average weights (FI_t)	Output change 2005-2019
	%		
Criminal cases (CrC_t)	19.50	17.82	-24
Civil cases (CL_t)	27.19	29.10	0
Administrative cases (AD_t)	10.67	5.52	-33
Immigration cases (IM_t)	5.31	5.98	-48
Cantonal cases (Ca_t)	17.22	19.44	-12
Tax cases (TC_t)	3.18	3.01	178
Civil law higher appeal (CrA_t)	6.51	7.35	-1
Criminal law higher appeal (CLA_t)	6.04	6.82	-13
Tax cases higher appeal (TCA_t)	1.51	1.71	-72
Central Board of appeals (BA_t)	2.87	3.24	-11

Table 4.1.5.2 as well as Figure 4.1.5.3 show that overall the courts experienced a decline in productivity of 15 percent if the appeals are counted at both courts and a decline of 14 percent if the appeals are excluded from output calculations for first instance courts between 2005 and 2019. The difference in TFP between the two approaches of handling appeals is very small which suggest that according to this quality adjustment approach no large changes in quality are evident. The overall decline in TFP can possibly be attributed to the fact that almost all case categories saw a decrease in volumes over the examined period (Table 4.1.5.2). Various reasons contributed to the decrease in outputs. The courts and the overall judiciary institution ascribe the output decline partly to an insufficient number of available judges (NOS, 2020).

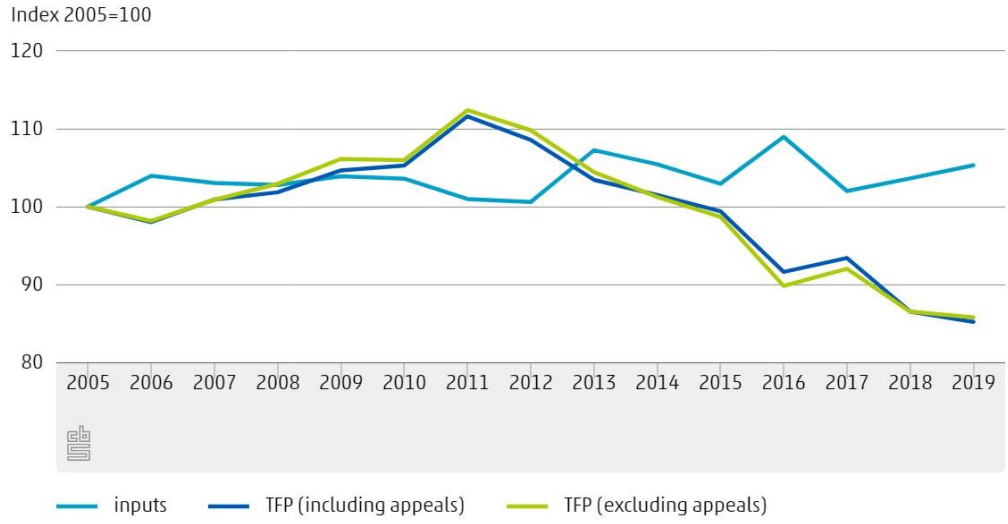
Table 4.1.5.2. Changes in inputs, output indicators and TFP of the courts, 2005-2019

	All cases counted		Excluding appeal cases in first instance	
Δ real inputs	Δ outputs	Δ TFP	Δ outputs	Δ TFP
%				
5	-10	-15	-10	-14

Additionally, the courts have to face a changing complexity of cases and court hearings. These require more witnesses, interpreters and other experts being part of the cases which increases handling times (NOS, 2020). Furthermore, the judges aim to increasingly take into consideration the defendant's personal background and underlying problems which increases the time it takes to handle cases. Ideally, a productivity analysis would incorporate the development of changing case complexity. However, for this more specific data is needed. Van der Ploeg and de Wit (2015) propose a possible approach to make inferences about developments of

case complexity by conducting dossier research and using different indicators such as the thickness of the case dossiers, the number of hearings per case or the number of expert reports. The authors find inconclusive evidence concerning the development of case complexity between 2008 and 2014. While for some case categories such as divorce cases, the complexity seems to have increased, for others such as trade cases at cantonal courts it decreased (Van der Ploeg and de Wit, 2015). This type of research requires very detailed information which goes beyond the scope of this paper.

Figure 4.1.5.3. Development of inputs ant TFP of the courts



In terms of output, the activity indicators seem to properly reflect outputs, the case categories are fairly homogenous and there is no indication that there are approaches where correcting for quality would lead to significantly different output and TFP estimates. Therefore, for the courts, it could be possible to provide reliable productivity estimates. However, the output indicators do not indicate changes in societal outcomes. This refers to the question of what courts want to achieve rather than what they actually do. Courts are supposed to provide an accurate application of the law in an impartial, just and efficient manner (Rechters voor Rechters, n.d.). Relevant aspects in terms of outcomes would thus be the degree of fairness of trials (i.e. does everyone receive the same treatment in court) or the degree of accessibility of the justice system both in terms of financial accessibility (i.e. is a fee necessary) but also whether citizens know about their options to access the justice system (OECD, 2015; Hauser, 2017). It is not clear whether the work of the courts in past years has contributed to a greater degree of fairness or accessibility. Hence, while in terms of output activities rather reliable productivity conclusions can be reached, this is not the case for societal outcomes.

4.2 The public prosecution service

4.2.1 Overview

The Openbaar Ministerie (OM) is the Dutch public prosecution service. The OM aims to provide a safe and just society by enforcing criminal law (Openbaar Ministerie, 2020a). As such it is the institution responsible to ensure that criminal offences are detected and prosecuted. In The Netherlands, the OM is the only institution that can bring suspects in a criminal court. Note that the OM only deals with criminal law and not with civil offences. The funding for the OM comes from national budgets (Blank and van Heezink, 2017). According to its website, the OM determines the following as its main tasks (Openbaar Ministerie, 2020a):

- Prosecuting criminal offences and bringing suspects in a criminal court
- Dealing with criminal offences without the intervention of a judge
- Leading the police in detecting criminal offences

Besides these well-known tasks, the OM also takes care of two other tasks, which it terms 'duties' (Openbaar Ministerie, 2020a):

- A prosecutor can ask a judge to impose mandatory care on persons if they are not able to care for themselves or possibly poses a danger for themselves or others (e.g. due to a mental disorder or a psychiatric disability)
- If a person has died a death not due to natural causes, the OM has to grant permission for the body to be buried

Although the OM is a national institution, it is organized in ten regions, termed arrondissementen ('districts') (Openbaar Ministerie, 2020a). These ten regions are consistent with the ten regional units of the police. In each region, the OM has an office called the arrondissementsparket ('district office') which is associated with the district court in the respective region. In the case that a suspect or a public prosecutor wants to appeal against a decision taken in the district courts, there are four ressortsparketten ('higher appeal office') (Openbaar Ministerie, 2020a). These deal with the reassessment of the case and are linked to the four courts of appeal which hand down a new verdict. Besides, the OM also has a national office that handles cases concerning (inter)national criminal activity such as human- or drug trafficking and cooperates with national institutions from other countries if needed (Openbaar Ministerie, 2020a).

4.2.2 Inputs

The inputs for the OM are labour inputs, other current expenditure inputs and intermediate inputs. Similarly to the courts, data concerning the inputs cannot be obtained from the COFOG database because the available data is not sufficiently disaggregated. Therefore, this analysis will also derive information from the annual reports published by the OM. The utilized inputs are presented in Table 4.2.2.1. Unfortunately, there is no information concerning capital expenses. The aggregated real inputs show an increase of around 12 percent between 2013-2019. Appendix A elaborates on the aggregation of inputs and the incorporated Figure A.2 shows the evolution of inputs in more detail.

Table 4.2.2.1. OM's inputs, measurement, % change in (deflated) inputs and sources

Input[24]	Measurement	Change 2013-2019	Source
Intermediate Consumption	Million euro	-46%	Openbaar Ministerie (2014-2020b)
Employee remuneration	Million euro	+16%	Openbaar Ministerie (2014-2020b)
Other current expenditures	Million €	+30%	Openbaar Ministerie (2014-2020b)
Employees	FTE	+7%	Openbaar Ministerie (2014-2020b)
External labour costs	Million €	+96%	Openbaar Ministerie (2014-2020b)

4.2.3 Outputs

To quantify the outputs of the OM it seems reasonable to start from the tasks which the institution itself regards as its functions. The OM aims to secure a safe and just society. This can be achieved by ensuring that criminal offences are detected, prosecuted and criminal offenders are being charged for their crimes. One might be tempted to solely measure the total number of cases handled. Nevertheless, the OM also engages in other activities such as investigation activities and the "special duties" discussed above. Thus, solely the total number of cases handled does not adequately represent the output produced by the OM. Unfortunately, StatLine does not contain any data concerning the OM. All information used in this analysis has been retrieved from the annual reports (Openbaar Ministerie, 2014-2020b). Since data concerning the inputs are only available from 2013 onwards this analysis considers the period of 2013-2019. Table 4.2.3.1 provides a summary of the variables used in the analysis.

Table 4.2.3.1. OM's output indicators and sources

Output indicator	2013	2019	Source
Number of felony cases at court (x1000) ($FC_{,t}$)	111.7	95.2	Openbaar Ministerie (2014-2020b)
Number of administrative cases at court (x1000) (AC_t)	56.5	62.1	
Number of higher appeals at court (x1000) (HAC_t)	19.0	17.2	
Number of traffic fines appeals at court (x1000) (ATC_t)	50.3	34.3	
Number of felony cases handled without intervention of a judge (x1000) ($F_{,t}$)	115.2	104.3	
Number of administrative cases handled without intervention of a judge (x1000) (A_t)	47.9	61.7	
Number of felony cases handled without intervention of a judge (x1000) ($TF_{,t}$)	514.1	375.7	
Number of Investigations Rijksrecherche (RI_t)	109	110	
Other Duties (x 1000) (D_t)	20.7	32.25	Openbaar Ministerie (2020b)
	2015	**2019**	Openbaar Ministerie (2016-2020b)
Investigations (x1000) (I_t)	1.765	2.295	

Task 1 – Bringing suspects of criminal offences before a court

One of the main functions of the OM is to bring suspects of criminal offences in a court. To measure this dimension, the number of cases that are brought to trial can

24 The costs items available in the annual reports include labour costs, material costs which were classified as intermediate consumption, external labour costs and so-called programma-uitgaven which were classified as other current expenses.

be retrieved from the annual reports. It has to be distinguished between different types of cases: (1) misdrijfzaken ('felony cases'), (2) overtredingszaken ('administrative offences'), (3) hoger beroepszaken ('higher appeal cases') and (4) beroepen tegen verkeersboeten ('appeals against traffic fines'). A problematic aspect of deploying this output indicator is the possible time lag between the inflow of the case and the date of the trial. Cases might span across multiple years and in that instance, the year where the input of resources is recorded is not necessarily the year where the output is recorded. Although this challenge cannot be solved as that would require the investigation of every case individually, it should be noted.

The figures concerning this dimension of output should be adjusted for quality and afterwards aggregated. The adjustment proposed in this analysis makes use of the number of vrijspraken ('acquittals') and therefore disregards the approach using handling times of cases as used in other analyses such as Blank and van Heezink (2017) or suggested by ONS (2018). The reason is that this approach neglects the association between handling times and case complexity (Openbaar Ministerie, 2020b). However, it should be taken into consideration that the complexity of cases most likely increases over time. This assumption seems reasonable as the payment which courts receive per conducted trial is supposed to increase from 2020 onwards to compensate for increasing case complexity (De Rechtspraak, 2020). Higher case complexity can result in longer handling times of cases due to increased difficulty. Consequently, an increase in handling times not necessarily represents a decrease in the quality of the work of prosecutors. Nevertheless, the number of felony cases and administrative offences brought to trial will be adjusted for quality by correcting for the number of acquittals.

Here, the assumption is that suspects are taken to court only if there is sufficient evidence to support the view that they actually perpetrated the criminal offence they are charged with. Consequently, the premise is that if a prosecutor finds that there is not sufficient evidence against the suspect, he or she will decide to not prosecute and not bring that case before trial. The number of cases lost by the prosecutors can then be seen as an indicator of the quality of their work. This analysis will not incorporate the total volume of cases where the suspect was acquitted but instead use the ratio of acquittals to total cases brought to trial. Using the total volume could be misleading for example in the case that the total number of cases rises more steeply than the number of acquittals. Then, quality would increase instead of decrease as suggested by the rising total number of acquittals. Data concerning the number of acquittals is only available for the case categories of felony cases and administrative offences. The quality-adjusted measures for these two categories are calculated following equation (13) where *QFt* stands for the quality-adjusted measure of felony cases in year t, *FCt* is the number of felony cases brought before a judge and *ACFt* is the number of acquittals in felony cases in year t:

$$QF_t = FC_t * (1 - {ACF_t}/{FC_t}) \tag{13}$$

Similarly, the quality-adjusted measure of administrative offences QA_t can be calculated with the respective numbers. Generally, one would want to also adjust the other two case categories for quality. However, no acquittal figures are provided. As the annual reports provide detailed information about the two case categories of felonies and administrative offences but only very little detail about the other two no quality adjustment is applied for the two categories.

When wanting to aggregate the measures, one directly encounters a problem: how to weigh the different case categories? Cost figures for the different case types are not available and there is also no data concerning the average working hours needed to handle cases of the different types. Therefore, one might be tempted to weigh each category equally in the sense that each case category is weighted according to its share of cases in total cases as is assumed for option 1 in Figure 4.2.3.3. However, it seems reasonable to assume that a felony case or a higher appeal case takes more time and resources to handle than an administrative offence or an appeal against a traffic fine. Consequently, these two case categories should receive higher weights in the aggregation. As the weights cannot be based on cost figures or working hours, the weights will be based on the price received by the courts for handling the different case types. This draws on the assumptions that the compensation received by the courts is a reasonable proxy for the costs of the respective case category and that the valuation for the respective categories is approximately equal for the prosecutors and courts. Felony cases are proxied by the price received by courts for a criminal case and the weight of a higher appeal case is based on the compensation for a higher appeal case in criminal law. Administrative offences and appeals against traffic fines are both handled by cantonal judges and thus receive the same price. To derive the weights, the realised prices received by the courts for the different case categories are multiplied with the respective number of cases per case category and the share of each case categories' costs in total costs is calculated. The average prices, cases and resulting weights are shown in Table 4.2.3.2.

Table 4.2.3.2. Compensation, number and weight per case category

Case category	Average compensation for the court	Average case numbers per category	Average weight per category
	Euro	X1000	%
Quality-adjusted felony cases (QFC_t)	1 069.28	105.2	71.70
Quality-adjusted administrative offences (QAC_t)	159.12	57.3	5.84
Higher appeal cases (HA_t)	1 593.79	17.2	17.47
Appeals against traffic fines (ATC_t)	159.12	49.4	4.99

The activities of task 1 (*Tt*) can then be aggregated according to equation (14) with *QFt* and *QAt* being the quality-adjusted measures of felony cases and administrative offences and *HAt* and *ATt* denoting higher appeal cases and appeals against traffic fines in year t respectively:

$$T_t = QFC_t * wQFC_{t-1} + QAC_t * wQAC_{t-1} + HAC_t * wHAC_{t-1} + ATC_t * wATC_{t-1} \quad (14)$$

The evolution of the aggregated activities of task 1 is portrayed in Figure 4.2.3.3. It is visible that for both weighting options, the volume of cases brought in courts by the OM decreases.

Figure 4.2.3.3 Aggregated cases brought in court[25]

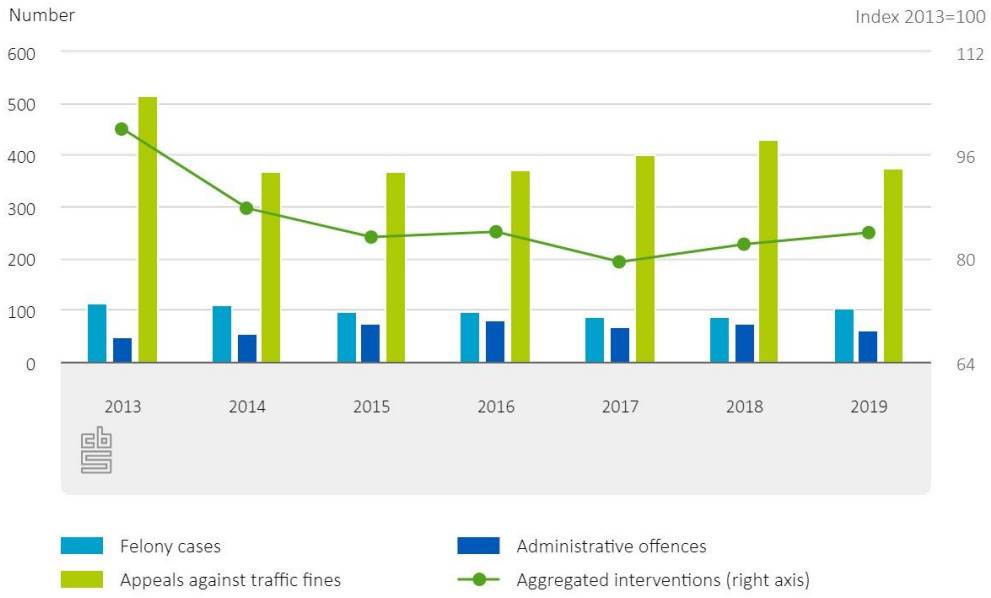
Task 2 – Dealing with criminal offences without the intervention of a judge

In The Netherlands, prosecutors have the authority to handle minor criminal cases and thus dispose of criminal offences without the intervention of a judge (Openbaar Ministerie, 2020a). While prosecutors cannot impose prison sentences, they can impose sanctions in the form of a fine, driving ban, community service or offer the offender a financial transaction (provided the suspect is found to be guilty of the criminal offence). This does, however, not apply to higher appeals cases. Additionally, the prosecutor also has the option to decide not to prosecute, which is called a sepot which occurs if the evidence is insufficient, the prosecutor assumes the evidence was not obtained in a lawful manner or the suspect is not identifiable (Openbaar Ministerie, 2020a). Similar to the output of task 1, the different case categories handled directly by the prosecutors have to be weighted differently. A similar technique as in the section before will be utilized to aggregate the outputs of task 2, as visible in equation (15). Note that here again a time lag may be present. Figure 4.2.3.4 shows the number and aggregated index of interventions.

$$OM_t = F_t * wF_{t-1} + A_t * wA_{t-1} + TF_t * wTF_{t-1} \quad (15)$$

25 Option 1: fixed weights means that case categories are weighted according to their share of cases in total cases. Option 2: different weights means that case categories are weighted according to the compensation received by the courts for the respective case category in total costs.

Figure 4.2.3.4. Number and index of interventions



Task 3 - Investigations

Leading the police in detecting criminal offences is task 3. While there is no exact data concerning the number of cases where the OM participates in investigations, there is information about the number of investigations into criminal organizations conducted by the district offices, the higher appeal offices and the national office. The caveat is that this data is only available from 2015 while generally 2013 is used as a base year in this analysis. Therefore, the number of investigations into criminal associations is discarded for now and further estimates of TFP regarding the period 2015-2019 including all investigations will be presented at the end of this analysis. Nevertheless, another part of the OM conducts investigations, the Rijksrecherche ('national criminal investigation department'). This body investigates behaviours that could harm the integrity of the government or suspected criminal offences conducted by other civil servants such as a prosecutor or policeman (Openbaar Ministerie, 2020a). Although it is part of the police, it is under the responsibility of the OM and thus the activities are considered in this analysis. The annual reports only provide very little information concerning the activities conducted by the Rijksrecherche. It is possible to observe the change in volume over time but there is no information about how successful the investigations are (Openbaar Ministerie, 2016-2020b). Here, "successful" is meant in the terms of investigations where it is found that either the suspect is guilty or not guilty based on sufficient evidence. Unsuccessful investigations would thus be those that are stopped and where no conclusion is reached. However, due to a lack of data, this cannot be incorporated into the analysis.

Other duties – Wet Bopz

In the overview, it is mentioned that besides its main tasks, the OM also has certain special tasks. Data concerning the number of times the OM has to grant permission for a body to be buried is not available online nor in the annual reports and consequently, it is not considered in this analysis. For the second special duty,

namely that public prosecutors ask judges to impose mandatory care on someone who is possibly endangering himself or others (from here on this type of activity will be termed Wet Bopz[26] activity due to the law it is based on), an approximation is published in the annual report of 2019 (Openbaar Ministerie, 2020b). It states that the OM had to engage in such activities 32,250 times in 2019 which was a large increase compared to 2014 when the OM carried out this duty only 24,000 times. This shows an increase of 34.38 percent. Under the very strong assumption that the increase in Wet Bopz activities is linear, it is possible to calculate the slope *m* of the curve with *y2* and *y1* being the number of the Wet Bopz activities 32,250 and 24,000 respectively and *x2* and *x1* being the years 2019 and 2014 respectively.

$$m = {}^{y_2 - y_1}/_{x_2 - x_1} \tag{16}$$

The result suggests that the volume of Wet Bopz activities increases by 1,650 cases per year. Consequently, it is possible to estimate the number of Wet Bopz cases for the years between 2014 and 2019 but also for 2013 and incorporate it in the analysis. There is only limited information available about the prosecutor's Wet Bopz activities, hence a quality adjustment is rather difficult.

4.2.4 TFP analysis

To calculate the productivity of the OM, the different dimensions of output first have to be aggregated (equation (17)). In an ideal case, the weights in this aggregation would be based on the marginal valuation of the different outputs by consumers. However, this information is not available. It is also not possible to attribute cost shares to the different activities and information about input usage such as working hours per activity is also lacking. Therefore, the weights have to be based on the author's assumptions. It seems reasonable to make the premise that task 1 (bringing suspects before a court) and task 2 (handling cases without the intervention of a judge) constitute the largest shares of the work of the OM. Investigations probably comprise only a minor part of production, especially because the investigations conducted by the different offices are excluded in this part of the analysis and only Rijksrecherche investigations are included. The following TFP calculations are purely based on assumptions concerning the weights and therefor four different options of aggregation are proposed to calculate the outputs (Table 10). For option 1, rather extreme weights are chosen meaning that task 1 and 2 receive really large weights and the other two tasks receive small weights. The other options will portray the productivity of the OM if less extreme weight choices are made. The weights are displayed in Table 4.2.4.2. Once, the aggregated output measure is calculated, TFP can be derived by dividing the aggregated output measure by the aggregated input measure. Table 4.2.4.1 presents the changes in inputs and outputs while Table 4.2.4.2 and Figure 4.2.4.3 portray the TFP estimates for all four weighting options.

26 This refers tot he Wet bijzondere opnemingen in psychiatrische ziekenhuizen ('Special Admissions to Psychiatric Hospitals Act'). It should be noted that from January 1st 2020 onwards this law is obsolete, it was replaced by the Wet verplichte geestelijke gezondheidszorg and the Wet Zorg en dwang (Inspectie Gezondheidszorg en Jeugd, n.d.).

$$O_t = T_t * wT + OM_t * wOM + RI_t * wI + D_t * wD \quad (17)$$

Figure 4.2.4.3 and Table 4.2.4.2 reveal that in all presented scenarios, the OM experiences a decrease in overall TFP. Depending on which option of weighting is chosen, the TFP estimates differ significantly. Option 4 shows an overall decrease in productivity of 11 percent while using option 1, the productivity decreased by 20 percent. Especially the two main outputs, the cases taken to trial and interventions without a judge, which presumably constitute the largest share of the prosecutors' work decreased in volume.

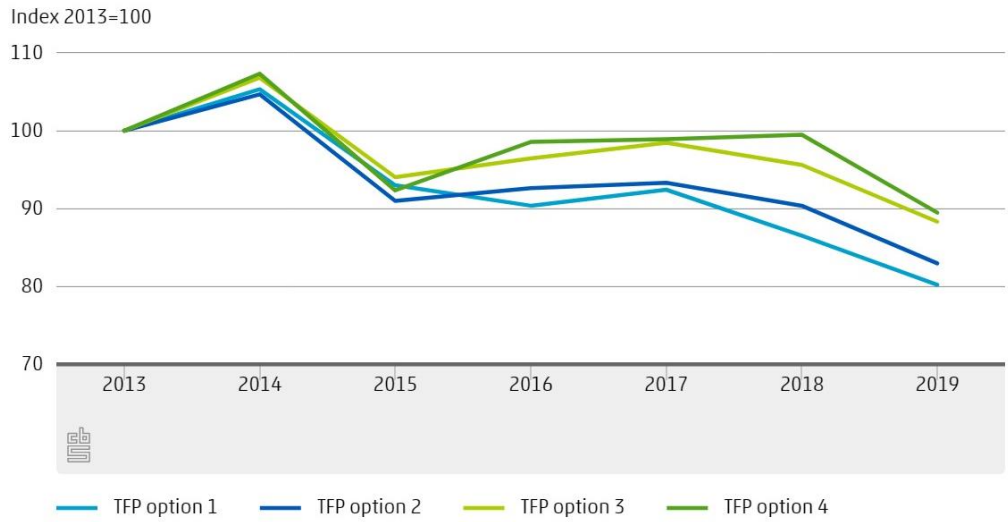
Table 4.2.4.1. Changes in OM's real inputs and output indicators 2013-2019

	Outputs			
Δ real inputs	Δ Task 1 - (T_t)	Δ Task 2 - (OM_t)	Δ Task 3 - (RI_t)	Δ Task 4 - (D_t)
%				
12	-12	-16	1	44

Table 4.2.4.2 Possible weighting options for output aggregation and changes in TFP

Activity	Task 1 Trials (T)	Task 2 Interventions (OM)	Task 3 Rijksrecherche Investigations (RI)	Wet Bopz (D)	TFP change 2013-2019
	%				
Option 1	60	30	5	5	-20
Option 2	40	40	10	10	-17
Option 3	40	30	10	20	-12
Option 4	30	30	20	20	-11

Figure 4.2.4.3. Development of OM's TFP for different weighting options



The OM provides the following reasons for the decrease in output volume in its 2017 report: increasing complexity and cases become more labour-intensive which negatively impacts productivity (Openbaar Ministerie, 2018). The OM argues that prosecutors aim to increasingly consider the individual situation of suspects even in simple cases and not only investigate the case from a criminal perspective (Openbaar Ministerie, 2020b). In many cases, the criminal offence is associated with underlying issues such as mental health, debts or addiction. Therefore, the prosecutors have to increasingly collaborate with partners such as the addiction control centre to do the case justice (Openbaar Ministerie, 2020b). Additionally, the public prosecution service was confronted with insufficient capacities at courts meaning that the courts did not have sufficient personnel and sitting capacities to deal with the influx of trials (Openbaar Ministerie, 2020b). All these issues result in longer handling times, lower outflows of handled cases and consequently lower estimated productivity. This can only to a certain extent be offset by the strong increase in the volume of Wet Bopz activities. The public prosecution service is expected to deal with cases in more depth while its other duties, the Wet Bopz activities also increase. Hence, it is not necessarily surprising that the OM experiences a decrease in overall output. It should also be noted that changes in case complexity are not actively accounted for in this analysis due to a lack of proper indicators of case complexity.

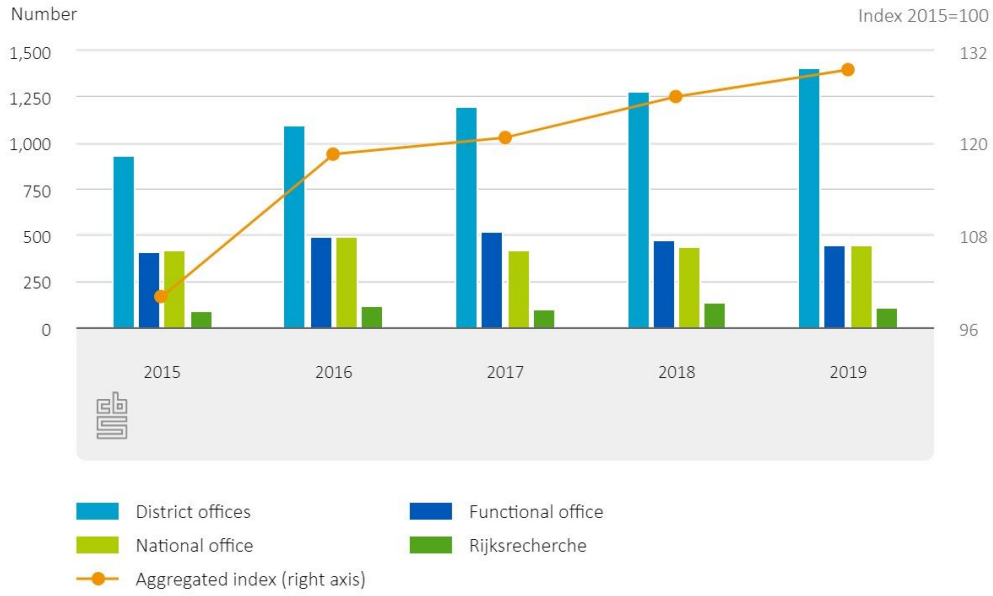
These TFP estimates are derived using many assumptions and despite the lack of important data. More consistent data concerning the Wet Bopz activities and information concerning the compensation of the prosecutors for cases of different case categories would be desirable to make results more reliable. If more granular data concerning the cost structure of the public prosecution service was available it would be possible to properly aggregate the different output dimensions and to estimate its productivity more accurately. In conclusion, productivity estimations for the public prosecution service require the application of numerous assumptions. However, greater data availability and information concerning the cost structure could benefit the productivity estimates and contribute to less sensitive results. Nevertheless, the contribution of the OM to societal outcomes in terms of a safe and just society would be of great interest besides the performance in terms of outputs. The OM is supposed to propose punishments that minimize the chance of reoffending as much as possible and as such reduce criminal behaviour of citizens and contribute to lower levels of crime (Openbaar Ministerie, 2020b). However, the exact contribution of the OM to the prevention of reoffending cannot be determined.

4.2.5 Analysis including investigations

In the analysis presented above, only Rijksrecherche investigations are included in the volume output measure of task 3 since other investigations data is only available from 2015 onwards. Therefore, this section will provide an analysis using similar assumptions as above, however for the period 2015-2019 and thus including data concerning the investigations conducted by the public prosecution service. The OM investigates suspected subversion in the form of among others drug trafficking,

human trafficking but also environmental crimes, cybercrime or money laundering. The investigations conducted by all offices concern complex crimes which is why it does not seem appropriate to apply different weights. Therefore, the output volume of task 3 is calculated by adding the number of Rijksrecherche investigations and the number of investigations conducted by the different public prosecution offices. Figure 4.2.5.1 shows the development of investigations conducted by the different offices and the volume index. It is evident that the output volume is much larger when including all investigations and also the aggregated index increases by 30 percent (Table 4.2.5.2) in contrast to only 1 percent in the previous analysis (Table 4.2.4.1).

Figure 4.2.5.1. Number and index of investigations including investigations of OM offices



In contrast to the previous analysis, data for all three main tasks of the OM is available. Consequently, all three main tasks will receive equal weights, namely 30 percent each, while the special task (Wet Bopz) will only receive a weight of 10 percent in aggregation. Nevertheless, to provide better comparability to the results of the previous analysis, the output and consequently TFP is also calculated using weighting option 1 from the previous analysis.

When calculating TFP including the investigations with weighting option 1 from the previous analysis, the overall productivity index still shows a decline, albeit only by 16 percent (Table 4.2.5.2). In contrast, weighting option 1 led to a decline of productivity of 20 percent previously. This difference can be attributed to the increase in the output of Task 2 (OM). Notably, between 2013 and 2019 the output of task 2 decreases by 16 percent while the output of this task increases by 1 percent between 2015 and 2019. When using the weighting option that assigns equal weights to the 3 main tasks of the OM, the decrease in productivity is equal to only 5 percent. This smaller productivity decline can be attributed to the large

increase in investigation activities between 2015 and 2019 (+ 30 percent) as the OM puts a focus on fighting organised crime. In combination with a larger weight attributed to this activity, the increase in investigations leads to a smaller decrease in TFP.

Table 4.2.5.2. Changes in OM's inputs, output indicators and TFP, 2015-2019

	%
Δ real inputs	13
Outputs	
Δ Task 1 – (T_t)	-13
Δ Task 2 – (OM_t)	1
Δ Task 3 – (I_t)	30
Δ Task 4 – (D_t)	26
TFP estimates	
Weighting option 1 from previous analysis	- 16
Weighting option 2 – weight T1, T2, T3 = 30%, weight Wet Bopz = 10%	- 5

4.3 The police

4.3.1 Overview

The Dutch Politie ('police') is one national police force. Its motto is "The police are always there, for a safer Netherlands" as well as "waakzaam en dienstbaar" which can be translated as "watchful and ready to serve" (Politie, n.d.a). The police aim to provide a safe and pleasant environment for Dutch citizens to live in, ensures that the Dutch democracy is protected and that the law is enforced. The organization of the police has seen a large reconfiguration based on the Police Act 2012, which came into force in 2013 (Terpstra and Fyfe, 2019). The consequence was a change from a regionalised form with 25 regional corps and the Korps landelijke politiediensten ('national police corps') to the national police force present today (Urlings, Blank and Niaounakis, 2014). The force is organized in ten regional units and one central unit which handles more specialised duties and acts as a support for the regional units (Politie, n.d.a). Besides, the Police Service Centre is part of the national police force. It takes care of all operational management tasks such as finances or human resource management (Government of the Netherlands, n.d.). The authority over each of the ten regional units lies with a Chief Constable meaning that the power of authority is still focused at the local level, however, decisions concerning police resources are now taken at the national level (Terpstra and Fyfe, 2019). The Ministry of Justice and Safety is responsible for the functioning of the national police force (Stroucken, 2016).

4.3.2 Inputs

The information concerning the inputs of the police is available from the COFOG database. The police are part of the first-level category of Public Order and Safety. In addition to the cost figures provided in the COFOG database, the cost of volunteers, as well as information concerning the labour inputs in FTE, is retrieved from the annual police reports. The inputs used in the analysis as well as their

sources are listed in Table 4.3.2.1. Note that the assumption that inputs are remunerated according to their marginal productivity is problematic in the case of volunteers. The nature of volunteer work incorporates that volunteers receive only a low or no monetary compensation for their work. However, the cost share of volunteers in total input costs is equal to only 0.07 percent on average. As the costs of volunteers constitute only such a minor part of inputs, this problematic aspect can be disregarded. More detailed information concerning the aggregation and figures displaying the evolution of inputs are supplied in Appendix A (figures A.3 and A.4).

Table 4.3.2.1. Police's inputs, measurements, change (deflated) inputs and sources

Input	Measurement	Change 2013-2019	Source
Intermediate Consumption	Million euro	-5%	CBS (2020d)
Employee remuneration	Million euro	+31%	CBS (2020d)
Other current expenditures	Million euro	-41%	CBS (2020d)
Other capital expenditures	Million euro	+35%	CBS (2020d)
Cost of Volunteers	Million euro	-10%	Politie (2015-2020)
Employees	FTE	-0.5%	Politie (2015-2020)

4.3.3 Outputs

Measuring the outputs of the police is rather difficult since the conceptualisation of output is not clear. What exactly is the output of the police? Previous studies have used a range of different output indicators (Drake and Simper, 2003). Generally, it is assumed that the police is responsible for providing a safe environment and that its activities have a large impact on the degree of criminal activity, public order and safety in the country (Zouridis, 2015). Nevertheless, following this strain of thought disregards the fact that criminal activity and the extent of public order and safety depend on several factors and the police are only one of them (Zouridis, 2015). Thus, the author urges the reader to note that crime is a complex concept that is subject to many external factors. Hence, it can be difficult to directly link changes in the extent of crime to police work.

When attempting to conceptualize the output produced by the police, it is worth considering what the Dutch police itself regards as its tasks: the police states that its tasks are to ensure safety for its citizens, prevent and detect criminal activity, maintain public order and provide help where needed (Politie, n.d.b.). Additionally, it is important to take into consideration the valuation by citizens. According to Lind and Lipsky (1971), citizens do not derive utility directly from the activities conducted by the police but from the outcome, its outputs contribute to. Therefore, in theory, the aspect of interest is here the change in outcomes to the extent that these changes can be attributed to police activity[27]. The outcome which is valued by citizens is crime reduction and having a safe living environment. The former can be achieved through two paths: by detecting and fighting crime and consequently charging criminals as well as engaging in activities that aim to prevent

27 Cutler et al. (2021) present how to introduce an approach that measures changes in outcome to make inferences about productivity changes in medical care.

crime from happening. Additionally, providing a safe environment for citizens entails providing community support, responding to emergencies and maintaining public order. Therefore, the police's tasks can be divided into five larger fields (1) crime control (2) crime prevention (3) emergency management (4) maintaining public order (5) community support. The problem with the field of community safety and support is that it is mainly concerned with outcomes. There is no information concerning direct output produced by the police or how much of the change in outcomes can be attributed to the police. Thus, this dimension will be disregarded in the productivity analysis but developments of outcomes will be presented in the discussion at the end. It should also be noted that the dimensions of crime control and crime prevention are to some extent associated. Therefore, it is of major significance to include the dimension of prevention in the analysis as otherwise a potential increase in registered crimes implies an increase in productivity, which is counterintuitive to the outcome valued by citizens.

Most of the information concerning the output indicators in this analysis can be retrieved from StatLine but certain information such as the number of patrol officers or convictions have to be obtained from different sources. Table 4.3.3.1 portrays the output indicators used and their sources.

Table 4.3.3.1. Police's output indicators and source

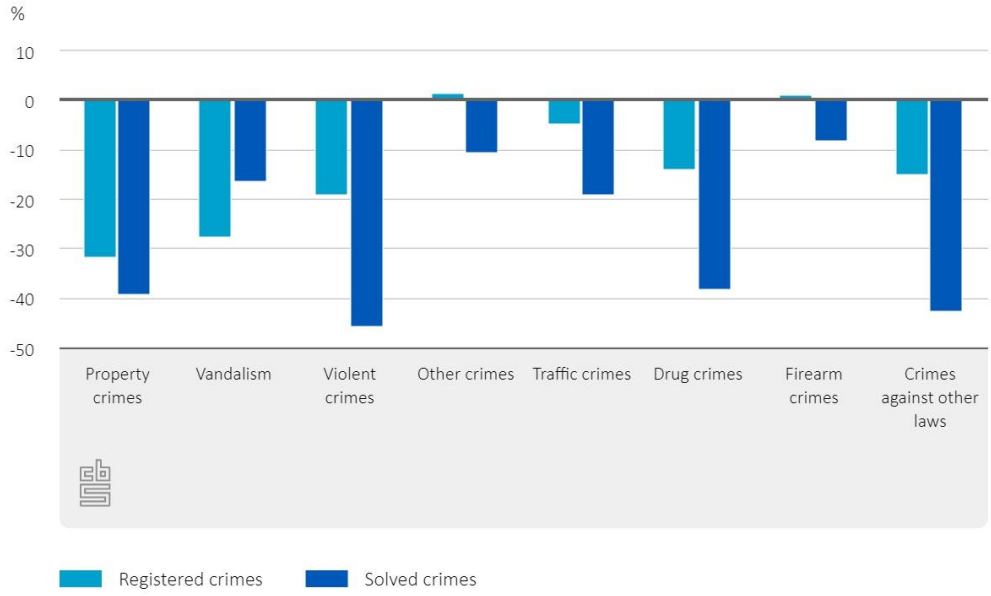
Output indicator	2013	2019	Source
Number of registered crimes (RCk,t)	1 105 565	820 505	CBS (2021c)
Number of solved crimes (SCk,t)	289 665	229 380	CBS (2021d)
Number of investigations (CIt)	950	1 522	Politie (2015-2020), OM Jaarbericht, (2013, 2014)
Number of actionable emergency calls (EMt)	1 647 590	1 711 000	Politie (2015-2020)
Number of patrol officers (WAt)	3 456	3 542	Politie (2015-2020), Hulpverlening.nl (2014)
Dutch population (mln) (POPt)	16.8	17.3	The World Bank (2020)
	2013	2017	
Number of convictions in first instance (Ck,t)	92 900	82 700	Trends in the Netherlands (CBS, 2019a)

Crime control

The analysis of outputs will start by observing the activities conducted to control crime by detecting and fighting criminal activity. Here, several different indicators are presented. Other studies such as Blank and van Heezink (2017) or the Demonstration Project (Home Office, 2001) use the number of registered crimes. The intuition behind using this indicator is the assumption that officers work a case, even if they do not manage to solve it. This implies that the police still applies resources to register the case, for example, by taking the victim's statement and investigating possible leads. Thus, the registered cases can be regarded as outputs. However, presumably, citizens do not only expect the police to register crimes but mainly to solve crimes. Therefore, the second indicator proposed is the number of solved crimes, so the number of clearances. Solving the crime and finding the suspect is a premise to charging the criminal and increasing the degree of safety.

Hence, the number of clearances better describes the outcome valued by citizens (Blank and van Heezink, 2017). Anyways, Blank and van Heezink (2017) argue that both indicators portray a strong correlation and the productivity estimates do not differ much depending on the chosen indicator. This is confirmed in the TFP analysis in this study. It is noteworthy that for both indicators, registered and solved crimes, the measure of interest is not solely the total number of crimes but the number of crimes per crime category. The analysis incorporates eight different crime categories which will be more relevant when it comes to aggregating the different categories into a single index. Figure 4.3.3.2 shows that except for other crimes and firearm crimes all crime categories saw a decrease in the number of registered as well as solved crimes from 2013 to 2019.

Figure 4.3.3.2. Change in registered and solved crimes 2013-2019



Although the output indicator of solved crimes comes closer to the outcome valued by citizens, this indicator does not give any indication of whether offenders are called to account for their actions. It might be that the police has the impression that the case is solved and a suspect is found but if the evidence is insufficient and the public prosecution service, as well as the courts, cannot confirm this suspicion, the crime cannot completely be considered “cleared” (Moore and Braga, 2003). To achieve the desired outcome of increased safety, criminal offenders have to not only be found but also charged. Therefore, the third indicator proposed is the number of criminal convictions at courts of first instances available from Trends in the Netherlands published by CBS (2019a). This data distinguishes the crimes in categories similar to the ones provided for registered and solved crimes on StatLine which makes it more comparable than data obtained from the public prosecution service. Unfortunately, the data is only available until 2017 which implies that the resulting TFP calculations only cover the period 2013-2017. Note that this indicator is also not without flaws because the number of convictions not only depends on the work of the police but also the work of prosecutors and courts (Moore and

Braga, 2003). The standard of proof, a court requires to find someone guilty might be higher than the standard of the police to make an arrest[28]. Nevertheless, this indicator to some extent provides information about the outputs of the police and also the quality of the outputs. Moore and Braga (2003) present evidence that increased quality of investigations is reflected in increasing conviction numbers. Similarly to the output series of registered and solved crimes, the number of convictions decreased for each crime category.

An important aspect is the weighting of different types of crimes in aggregation. The annual reports do not provide information about the costs of different crime categories, the amount of resources deployed or the value to citizens of handling different crimes. Nevertheless, it seems to be a reasonable assumption that solving crimes of different categories will have different values to citizens. For example, solving a violent crime is probably valued higher and requires more resources than solving a bike theft (Pritchard, 2002). Thus, the former category should receive a higher weight in aggregation as resource-intensive crimes should be seen as generating more output than low-resource crimes (Pritchard, 2002). This analysis proposes two weighting structures for the different crime categories. The first option bases the weights on the share of cases of each category in total cases/convictions. This implies that categories with higher case/conviction numbers receive higher weights than categories with lower numbers.

The second option adjusts the weights for the social costs of crime to society as presented in Heeks et al. (2018). Heeks et al. (2018) calculated the costs that arise as a consequence of crime by estimating the monetary value of stolen property, the resulting physical and emotional harm, the lost outputs and the costs to health and victim services. As their study focuses on the British context, the assumption is that the presented cost structure is also applicable to The Netherlands. It should be noted that the social cost categories presented in Heeks et al. (2018) are not entirely congruent with the crime categories in StatLine. For example, the cost categories of robbery, domestic burglary, theft of vehicle and theft from vehicle can all be classified as property crimes. In this case, the average of the respective social costs is taken to represent the social cost caused by property crimes (ASC_t^P). Table E.1 (Appendix E) provides an overview of the categorisation and the respective social costs. The social costs per category are multiplied by the number of cases per category to arrive at the total social cost caused by the respective categories (equation (18)). The social costs of crime are only available for the "traditional" crime categories of property crimes, vandalism offences and violent crimes. However, these three categories make up at least 80 percent of registered crimes and around 70 percent of solved crimes. For the other five categories, social costs are not available, therefore these will receive the same weights as in the weighting option based on case numbers. As the weights are known, it is possible to calculate hypothetical social costs for the remaining categories. Thus, the total social costs of all crimes (TSC_t) can be derived and the social cost adjusted weights (w_t^j) are the

28 The court has to proof a suspect's guilt "beyond a reasonable doubt" while the police may only require "reasonable suspicion" to make an arrest (Moore and Braga, 2003).

share of the social costs caused by each crime category ($ASC_t^i * c_t^i$ in equation (18)) in total social costs caused by all crimes. The case of homicides that are classified as violent crimes deserves special treatment. Homicides are by far causing the highest social costs. Therefore, taking the average of the social costs for violent crimes including the social costs of homicides would result in huge social costs of violent crimes and thus constitute an overstatement in the weighting structure. Calculating the weight of registered crimes (wRC_t^v), one can use the number of homicides per year under the assumption that every recorded homicide is registered by the police. Hence, it is possible to calculate the social costs caused by homicides and the costs caused by other violent crimes[29] and add these (equation (19)).

$$w_t^i = \left(ASC_t^i * c_t^i\right) \Big/ TSC_t \tag{18}$$

$$wRC_t^v = \left(H_t * SCH_t + (VC_t - H_t) * ASV_t\right) \Big/ TSC_t \tag{19}$$

In equation (19), H_t denotes the number of homicides, SCH_t denotes the social costs of homicides, VC_t is the number of violent crimes and ASV_t describes the average social cost for violent crimes, the subscript t denotes the year. However, for solved crimes and convictions, there is no data concerning the number of solved homicides or homicide cases with a conviction. Hence, to refrain from significantly overstating the social costs caused by the category of violent crimes, here the average social costs excluding the costs of homicides is applied. Table 4.3.3.3 shows the weighting structure resulting from the two approaches. It is noteworthy that even if these measurable output indicators are considered, the weighting attributed to the different types of criminal activities differ significantly. For example, the weight ascribed to property crimes ranges from 60 percent (when using the indicator of registered crimes) to 16 percent (when using solved crimes). This already points to the variation in TFP estimates that will be apparent in the next section.

As stressed before, quality adjustment is an important aspect of measuring output. The weighting of different crime categories can be regarded as an implicit quality adjustment. The intuition is that the weighting accounts for shifts in the case-mix. For example, an increase in solved vandalism crimes at the expense of violent crimes would overall constitute a decrease in quality. It should be noted though that this approach only considers shifts in the case-mix but not changes in complexity within a crime category. There is no accurate measurement for the latter. Due to the conceptual problems, it is not possible to implement explicit quality adjustment approaches.

29 The average social cost of violent crimes is derived by excluding the costs of homicides.

Table 4.3.3.3. Average weights for different types of criminal activity[30]

Average weights:	Registered crimes		Solved crimes		Convictions[31]	
	based on share in total case numbers	adjusted for social costs	based on share in total case numbers	adjusted for social costs	based on share in total convictions	adjusted for social costs
	%					
1 - Property crimes	60	40	34	16	37	18
2 - Vandalism	13	6	11	3	10	3
3 - Violent crimes	10	37	23	50	20	45
4 - Other crimes	1	1	3	3	17	17
5 - Traffic crimes	13	13	19	19	8	8
6 - Drug crimes	2	2	6	6	9	9
7 - Firearm crimes	1	1	2	2		
8 -Crimes against other laws	0.4	0.4	1	1		

Another possible indicator for the dimension of crime control is the number of investigated criminal organisations. These are entities that engage in organised crime that is dangerous to society. The increase in these investigations between 2013-2019 is significant. It should be considered that the police does not conduct these investigations alone but in cooperation with the public prosecution service. It is not clear what part of the output can be contributed to the work of the police, however, this analysis will assume that the part contribution of the police work remains constant over the years. Under this premise, the evolution of the police output in terms of investigations in criminal organizations is similar to the evolution of total investigations.

The significant increase of investigations should be regarded in a critical light. Korf, Luijk and de Meijer (2018) find that in earlier years, the registration of investigations into criminal organisations was not as standardized as it is now which means there could be an underestimation of the number of investigations and consequently an overestimation of the recorded growth. Additionally, to make investigations more manageable, strategies such as shortening investigations, having investigations go less in-depth or scale reductions have been applied (Korf, Luijk and de Meijer, 2018). Consequently, it is possible to conduct more investigations in a shorter time. It is thus important to know whether the increase in investigations was at the expense of quality. Unfortunately, information concerning the number of arrests made or suspects charged is not available publicly which is why quality adjustment for this indicator is not possible.

Crime prevention

Many methods estimating the productivity of the police are subject to a "criminality bias" (Zouridis et al., 2014). This bias constitutes that a large focus on the crime-fighting activities of police work is evident and other tasks of the police are

30 Note that the weights in each column always add up to 100%.

31 In the case of convictions, the categories property crimes, vandalism, violent crimes, traffic crimes, drug crimes and other crimes are available which is why there are no weights for categories 7 and 8.

neglected. To prevent this analysis from being subject to criminality bias, other dimensions of police work are incorporated as well. The police are not only responsible for investigating past crimes but also for preventing future criminal offences from occurring. Naturally, it is very difficult to estimate the number of prevented crimes as this is completely hypothetical. Further, the number of prevented crimes is also dependent on socio-economic factors. Urlings, Blank and Nianouakis (2014), as well as Blank and van Heezink (2017), apply estimations of prevented crimes in their analyses. This approach is not chosen in the analysis at hand because it is not possible to accurately approximate the contribution of the police to the number of prevented crimes. Other factors such as prevention by private organizations or changes in society can also play a role. Nevertheless, the crime prevention dimension should not be disregarded as the decrease in output concerning registered crimes could be caused by increasing crime prevention.

Similarly, the decrease in solved crimes could be caused by a shift from solving crimes to crime prevention activities. While it is not possible to measure crime prevention activities that the police actively engages in, another possible indicator is the presence of the police. By assuming that the presence of the police to some extent negatively influences the emergence of criminal activity, crime prevention can be assumed to increase if the number of patrol officers increases. Vollard and Hamed (2012) find that higher police numbers decrease violent and property crimes in the United Kingdom. It is assumed that this also holds for the Dutch context. Additionally, the Dutch police have set as a requirement to have one patrol officer per 5 000 inhabitants. Therefore, this analysis will incorporate the number of patrol officers per 5 000 inhabitants as an indicator for crime prevention. To compute Pt, the number of patrol officers (WA_t) is divided by the quotient of the Dutch population (POP_t) divided by 5 000 (equation (20)).

$$P_t = {WA_t}\Big/{\left(\frac{POP_t}{5000}\right)} \tag{20}$$

Emergency management

The emergency management (EM_t) of the police is concerned with acting when called for an emergency and lending a helping hand to whoever is in need. One would be tempted to apply the number of 112-emergency calls as a volume indicator. However, the number of calls is not of interest but the number of resulting police operations gives a better indication of police output as not every answered 112-call requires action. Therefore, the number of police operations in response to the placed 112-emergency calls is the desired volume measure but this data is not published. The data available gives information about the number of 112-emergency calls in the years 2016-2019. The share of 112-emergency calls that resulted in a police operation is equal to 58% for the year 2019 which is the only year data is available (Politie, 2020). Therefore, the premise is that the share of emergency calls resulting in a police operation remains constant over the examined period. As there is no data for the years 2013-2015, it is assumed that the number of emergency calls was similar to 2016. Ideally, more granular data concerning this

aspect as well as data concerning the police operation resulting from emergency calls for all years would probably improve the analysis.

Maintaining public order

Another task of the Dutch police is to ensure the maintaining of public order. This includes being present at demonstrations but also events such as football games or festivals. Since the police do not publish any data concerning this dimension, defining a volume output series is challenging. According to Adang et al., (2014), festivals and other events have continuously increased in popularity and consequently also in number over the last few years. Especially large events with more than 5000 participants take place more often. Such events require a large police presence and thus use large policing capacities. Therefore, it seems like a reasonable assumption that the activities to maintain public order have increased in recent years. Nevertheless, the police have also stated that many of its activities conducted at such events could also be undertaken by other security personnel from the private sector. Requests such as the Wetvoorstel Politiekosten Evenementen which constitutes that organizers of large events should pay for the service provided by the police could provide disincentives to using police to maintain public order at events (Overheid.nl, n.d.). For now, the transfer of police costs to event organisers only applies to events in the context of soccer, however, the Vereniging van Evenementenmakers states that it could also be introduced for other large events (VVEM, n.d.). Consequently, it could also be the case that maintaining public order activities have either increased only at a small rate or even decreased. As there is no available information concerning this dimension, the TFP of the police will first be calculated including the other 3 dimensions. Afterwards, to investigate how each of the previously described possible developments of the maintaining public order dimension would influence productivity, this will be explored employing a sensitivity analysis. Three scenarios will be investigated: one with constant maintaining order activities, one with an annual increase of 5% and one with a decrease of 5% annually. The next section provides a more elaborate discussion of this aspect.

4.3.4 TFP analysis

To arrive at TFP it is necessary to divide the output measures by the input measures. Table 4.3.4.1 portrays the changes in real input and the different output measures. This table already hints at the development of TFP in the police.

Table 4.3.4.1. Changes in police's inputs, output indicators and TFP

Factor	Weighing type or scenario	Change 2013-2019
		%
Δ real inputs		2
Outputs		
Δ registered crimes	Case numbers	- 25
	Social costs	- 21
Δ solved crimes	Case numbers	- 20
	Social costs	- 23
Δ convictions*	Case numbers	- 11
	Social costs	- 11
Δ Investigations into criminal organizations		60
Δ Prevention		- 1
Δ Emergency management		4
Δ Maintaining public order	Scenario 1: 0%	0
	Scenario 2: 5%	34
	Scenario 3: -5%	- 26

* due to data availability, the period is 2013-2017

In Table 4.3.4.1 it is visible that the real inputs have seen a small increase over the examined period while most output indicators have seen a rather large decrease. This already points in the direction of a decrease in TFP in the police sector. Nevertheless, the combination and weighting of the different police tasks can result in variations of estimates for TFP. Table 4.3.4.2 presents the TFP for varying combinations of the different dimensions of police output.

Panel A in table 4.3.4.2 shows TFP estimates for "crime control TFP". Here, TFP is calculated using each of the four crime control indicators. The output estimates are obtained by using the social cost adjusted weights as these provide more accurate estimates than the weighting according to shares in case numbers. The different crime categories are aggregated according to their previous year's cost shares in social costs (in a similar manner as equation (14)). The three TFP estimates derived using the "classic" crime control indicators are negative. The result for the TFP calculation using the number of investigations shows a large increase in TFP. Naturally, basing police productivity only on the output concerning investigations is unreasonable but as this task forms part of police output it is included in this analysis. As no information concerning the inputs devoted to these specific tasks is available, it is not possible to derive more detailed conclusions. However, it could be assumed that such a significant increase in investigations is associated with a large increase in inputs devoted to that task. Hence, potentially investigation inputs showed a larger increase compared to the increase of total real inputs which if accounted for would adjust the productivity measure downward. Anyway, the productivity estimate should be observed carefully due to the caveats mentioned previously, for example, that the registration of such investigations was not always standardized (Korf, Luijk and de Meijer, 2018). Basing TFP on the three other (more

reasonable) crime control indicators leads to the conclusion that productivity is likely to have decreased over the examined period.

Table 4.3.4.2. TFP estimates for different scenarios for the police, index 2013=100

Output series			TFP estimates						
			2013	2014	2015	2016	2017	2018	2019
Panel A "Crime Control TFP"									
Registered crimes			100	96	92	88	80	77	77
Solved crimes			100	96	88	83	76	74	75
Convictions			100	100	104	92	91		
Investigations			100	123	129	149	147	149	157
Panel B Combinations with 2 dimensions									
Solved crimes and		W_{Sc} = 25%	100	95	90	87	92	91	94
Prevention		W_{Sc} = 50%	100	94	88	85	86	85	88
		W_{Sc} = 75%	100	94	87	82	80	79	82
Solved crimes and		W_{Sc} = 25%	100	98	96	95	88	94	97
Emergency management		W_{Sc} = 50%	100	96	93	90	83	87	90
		W_{Sc} = 75%	100	95	89	85	78	80	83
Panel C Combinations with 3 dimensions[32]									
Solved	Option 1	W_{Sc} = 33%							
crimes,		W_{P}= 33%							
Prevention		W_{Em}= 33%	100	99	96	93	91	92	92
and	Option 2	W_{Sc} = 50%							
Emergency		W_{P}= 25%							
management		W_{Em}= 25%	100	98	94	90	87	87	88
	Option 3	W_{Sc} = 15%							
		W_{P}= 42.5%							
		W_{Em}= 42.5%	100	100	98	96	95	96	96

However, this TFP analysis also includes other output indicators that provide information about the output concerning other dimensions of police work. From this point onwards, the indicator of solved crimes is used to reflect the dimension of crime control. The reason is that solved crimes more closely reflect an outcome that is valued by customers than registered crimes but is not as dependent on the work of other institutions as the number of convictions. Again, the output series using the social costs adjusted weighting is applied. Panel B of table 4.3.4.2 presents TFP estimates if the dimension of crime control is combined either with the dimension of prevention activities or emergency management. To estimate TFP, the dimension of crime control has to be aggregated with each of the two other dimensions to arrive at a single output index. Ideally, the weights in this aggregation would be based on the marginal valuation of the different outputs by consumers or the resource usage based on a cost structure. Since this information is not available different weight combinations are explored. Assuming that it is unrealistic that the two non-crime control dimensions make up more than 75 percent respectively, the TFP decrease ranges from only 3 percent (Crime Control and Emergency Management) to 18 percent (Crime Control and Prevention). This shows that even

32 W_{sc} = weight of solved crimes, W_{Em} =weight of the emergency response dimension and W_{P} = weight prevention.

when going beyond the crime-fighting activities of the police and including other dimensions as well, the TFP estimates decrease which is in line with the results obtained by Urlings (2012) and Blank and van Heezink (2017). The increase in emergency management output can only counteract the strong decrease in crime control output and cause an increase in TFP if it receives a weight of 90 percent or larger in aggregation which seems a rather unreasonable premise.

Nevertheless, it is interesting to observe changes in productivity when all three dimensions are incorporated in the output series. Again, the weighting is a problematic aspect that is subject to assumptions and arbitrariness. The weighting options shown in Panel C in table 4.3.4.2 are supposed to test the sensitivity of the analysis.

First, the TFP is calculated when all dimensions receive similar weights in the outputs aggregation. Option 2 assumes that crime control, so the traditional tasks of the police and the dimension that is mostly investigated in studies concerning police productivity, constitutes 50 percent of police output. The last weighting option is based on de Koning (2017) who states that police officers on the street only spend around 15 percent of their time solving crimes and mainly focus on the other dimensions of police work. Even if this refers to street police officers only, this scenario is also explored. In options 2 and 3, the prevention- and emergency management dimension receive equal weights as there is no reason to assume that consumer valuations or resources deployed would differ between the two.

Depending on the chosen weighting option for outputs, the calculated productivity of the police decreased by 8 percent (option 1), 12 percent (option 2) or by only 4 percent (option 3). This large variation is due to the extent to which the dimension of crime control receives a large weight in aggregation. On the two other dimensions, the police output only experienced a decrease of 1 percent or even an increase of 4 percent. Hence, when crime control receives a weight of 50 percent, the significant decrease in output of this dimension results in a large decrease in productivity (option 2). According to the police, the decrease in activities connected to crime control is due to a shift from the focus on the handling of crimes to a focus on the other dimensions. Possibly, the increase in the output of the emergency response dimension was at the expense of crime control as capacities are already occupied with the other tasks. However, this cannot be stated with certainty without more detailed data.

The analysis presented in table 4.3.4.2 disregards the dimension of police output that is concerned with maintaining the public order at events and demonstrations. To account for this part of police work, a sensitivity analysis for the three possible scenarios is conducted (constant output, +5 percent annually, -5 percent annually). The output series of this dimension shows large variations which reflect in the aggregated output measure. Depending on which weighting option is chosen to aggregate the different dimensions of police work, the change in aggregated output can range from an increase of 14 percent to a decrease of 16 percent. Table 4.3.4.3 shows the applied weighting structure.

Table 4.3.4.3. Different weighting options for aggregation of police output dimensions

Weighting option	Crime Control	Crime Prevention	Emergency Management	Maintaining Public Order
	%			
Option 1	25	25	25	25
Option 2	50	20	20	10
Option 3	15	17.5	17.5	50

Table 4.3.4.4 shows that TFP differs significantly depending on the chosen weighting option and scenario. Even within a scenario the variation can range from a decrease in TFP of 9 percent to an increase of 12 percent. Naturally, weighting option 3 where the dimension of maintaining public order receives a really large weight provides the most extreme results. The assumption of a weight of 50 percent in aggregation might be unreasonable. If this weighting option is excluded, the TFP estimates vary from a 2 percent increase (scenario 2, option 1) to a decrease of 15 percent (scenario 3, option 2).

Table 4.3.4.4. TFP estimates, sensitivity analysis for the police, index 2013=100

Output series		TFP estimates						
		2013	2014	2015	2016	2017	2018	2019
Scenario 1	Option 1	100	100	98	95	94	94	93
No change	Option 2	100	98	94	91	87	87	87
	Option 3	100	101	100	98	97	97	95
Scenario 2	Option 1	100	101	100	100	99	101	102
+5% annually	Option 2	100	99	95	92	90	90	91
	Option 3	100	104	105	106	108	111	112
Scenario 3	Option 1	100	99	95	92	89	88	87
-5% annually	Option 2	100	98	93	89	85	85	85
	Option 3	100	99	95	91	88	85	82

In conclusion, considering only the more reasonable scenarios (thus excluding the scenarios where only investigations are counted as outputs, maintaining public order receives a weight of 50 percent or that non-crime control indicators receive a weight of 75 percent) the majority of scenarios and combinations suggests a decrease in TFP. However, results are ambiguous as not all scenarios present a negative productivity trend.

Nevertheless, in the introduction to the analysis, it was stated that citizens do not directly value the output produced by the police but rather the outcomes in terms of increased safety or a decreased level of crime. Additionally, adjusting outputs for changes in outcomes can be considered explicit quality adjustment and could introduce explicit quality adjustment in the analysis of police productivity (Genet and Hayward, 2017). When observing outcome indicators such as the inverse of the

level of victimization, the level of prevented crimes, or the feeling of safety of Dutch citizens presented, all show an upwards trend and as such a positive change in outcomes (Appendix F, figure F.1). However, since it is not possible to find out to what extent these changes are attributable to the police such indicators cannot be included in the analysis. It would be desirable for future research to incorporate the direct influence of police work (e.g in the form of a solved crime or prevented crime) on these societal outcomes to obtain a more accurate picture of police productivity. For example, Chalfin and McCrary (2013), as well as Vollard and Hamed (2012), find that the police have a negative effect on violent and property crime. Similarly, Yeo (2019) finds that police negatively affect property crimes but he even provides more concrete results stating that an additional police officer prevents around 17 thefts a year.

To conclude, estimating the productivity of the police is a difficult venture. Therefore, it remains questionable whether analyses like the one presented above for the police should be conducted by official statistical agencies. The reliability of the analysis is inhibited by the numerous assumptions applied due to the lack of consistent data and the problem with conceptualizing the output of the police. It should also be taken into consideration that the work of the police is also greatly influenced by political and societal forces (de Jong, 2018). Consequently, difficulties arise because the focus of police work keeps shifting. During the examined period, the police were subject to three different security agendas (2013 and 2014, 2015-2018, 2019-2022) which had different goals and different focal aspects. For example, in 2012, the prosecution service and the police formulated the goal of doubling the number of investigated organisations of criminal activities compared to 2009 (Korf, Luijk and de Meijer, 2018). As is evident in the large increase of investigations into criminal organizations (60 percent) between 2013-2019, this goal has been reached. These shifts can portray the parts of police work that are being attended to or the types of crimes that receive more attention (de Jong, 2018). The annual reports further reflect these changing agendas. Each report provides different performance information and information available for one year is not published in the next report. Therefore, consistent data for output indicators is often not available. Anyways, the entire analysis hinges on numerous premises and depending on which options are chosen the productivity either decreases or increases. The conceptual problems that occur when attempting to measure the productivity of the police are so severe that it remains questionable if these can be circumvented even if more consistent data were available. Therefore, it might not be possible to arrive at reliable estimates.

5. Conclusion

The paper at hand aims to investigate the relevance and especially the measurement of public sector productivity. In this context, the Atkinson Review (2005) emphasizes the importance of direct output measures to accurately

estimate productivity in the public sector. The relevance of improved public sector productivity measurement lies in the important role of the public sector in the overall national economy. The present analysis showcases that the assumption that public sector productivity is equal to zero does not hold as the courts as well as the public prosecution service exhibit clear negative productivity trends. Consequently, the output=input convention results in misleading conclusions concerning the productivity of the economy and economic growth. Due to the previously identified measurement challenges, it is necessary to conceptualize the measurable outputs and investigate the production of outputs individually for each organization when estimating the productivity of a public sector institution. This paper aims to implement such an analysis for the Dutch courts, the public prosecution service and the police. The study finds that public sector productivity estimates are highly sensitive and productivity results differ significantly based on the chosen underlying assumptions.

For the courts, output measures can be obtained by observing the volumes of the different case categories handled. Due to the availability of prices, these output series can be aggregated using cost weights under the premise that the prices provide information about the valuation of the different case categories. These estimates suggest that courts have become less efficient which serves as an indication that policy improvements are required. In contrast, the case of the public prosecution service is more complex. As the public prosecution service engages in several activities that should all be considered to produce an output series that covers the entirety of the institution's output, the output is derived by aggregating the cases brought before trial, the cases solved by the OM, investigations and information concerning other special duties conducted by the OM. The lack of available data for certain outputs and the challenge of weighting the outputs series of the different activities in aggregation due to missing cost structures result in the application of many assumptions in the analysis. The most complex analysis concerns the police. The police aim to secure a safe living environment for Dutch citizens by engaging in crime control, prevention activities as well as providing emergency help. For the police, the discussion of quality is conceptually very complex. In particular, with regards to the changes in societal outcomes caused by the police but being unable to link changes in outcomes directly and unambiguously to police work inhibits arriving at reliable productivity estimates.

Statistical agencies such as CBS have the obligation to provide "reliable and coherent statistical information" (Business.gov.nl, n.d.) that are as precise as possible. The present study highlights that for the courts the developed methodology with the available data allows arriving at rather reliable productivity estimates using only a few assumptions (although the aspect of case complexity should not be disregarded). While the activity output indicators function as a satisfying proxy for outputs, the study, however, does not investigate changes in outcomes which could be of great interest to policymakers as citizens value not only outputs but also especially outcomes. This aspect could be further investigated in future studies. In contrast to the courts, the myriad of aspects that constitute the work of the public prosecution service and also especially the police and the

resulting application of assumptions impede achieving such reliability for these two institutions. A multitude of assumptions has to be deployed to arrive at results. As such deriving productivity estimates of the public prosecution service and especially the police can be considered academic work rather than statistical work. This does not imply that the productivity of the public sector, in general, and the public prosecution service cannot be estimated. Here the analysis of the productivity of the public prosecution service requires the following:

- more consistent and more detailed data concerning all investigation activities as well as the Wet Bopz activities
- information about either consumer valuation or the institution's cost structure for the aggregation of outputs

Improvements on these aspects are desirable and could contribute to more reliable results. In the case of the police, this analysis has shown that arriving at reliable productivity estimates including all dimensions of police work is difficult. The conceptual problems are too severe to be able to provide conclusive results. Here, further academic research into the impact of police outputs on outcomes could provide further insights. For example, investigating the extent to which a solved crime or conviction contributes to a lower level of crime or to what extent an increase in patrolling officers increases the degree of safety considering other external (societal) factors could advance the research into police productivity.

References

Adang, O., van Arkel, D., Esman, C., Vos, H., van Oorshot, W. and Stronks, S. 2014. Politie en evenementen: Feiten, ervaringen en goede werkwijze. The Hague: Boom Lemma uitgevers.

Annabel, J. 2019. Enhancing measures of non-market output in economic statistics: Progress paper. Australian Bureau of Statistics. Online. Available at: Enhancing measures of non-market output in economic statistics: Progress paper | Australian Bureau of Statistics (abs.gov.au) [Accessed 16 February 2021]

Atkinson, A. B. 2005. The Atkinson Review : Final report. Measurement of government output and productivity for the national accounts. Basingstoke: Palgrave MacMillan.

Blank, J. L. T. and van Heezik, A. A. S. 2017. Productiviteit van overheidsbeleid. Deel 3 – De nederlandse veiligheid en justities 1980-2014. Delft/The Hague: Eburon Academic Publishers.

Bouckaert, G. and Van de Walle, S. 2003. Comparing measures of citizen trust and user satisfaction as indicators of 'good governance': Difficulties in linking trust and satisfaction indicators. International Review of Administrative Sciences, 69(3): 329–344.

Business.gov.nl. n.d. Statistics Netherlands. Online. Available at: https://business.gov.nl/partners/statistics-netherlands/ [Accessed 23 May 2021]

Caplan, D. 2006. Non-market services – the view from national accounts. Paper prepared for the joint OECD/ONS/Government of Norway workshop Measurement of Non-market output in education and health. October 3rd-5th London.

Caves, D., Christensen, L. and Diewert, W. E. 1982. Multilateral comparisons of output, input, and productivity using superlative index numbers. The Economic Journal, 92(365): 73-86.

CBS. 2016. Supply and use tables and sector accounts. Online. Available at: https://www.cbs.nl/en-gb/custom/2016/35/supply-and-use-tables-and-sector-accounts [Accessed 1 April 2021]

CBS. 2019a. Trends in the Netherlands: Security and justice. Online. Available at: Figures - Society | Trends in the Netherlands 2019 - CBS [Accessed 14 May 2021]

CBS. 2019b. Supply- and use, input-output and sector accounts. Online. Available at: https://www.cbs.nl/en-gb/custom/2019/29/supply-and-use-input-output-and-sector-accounts [Accessed 2 April 2021]

CBS. 2020a. Bruto investeringen in vaste activa; bestemming, nationale rekeningen. Online. Available at: https://opendata.cbs.nl/#/CBS/nl/dataset/84090NED/table?dl=44645andts=1622299511217 [Accessed 2 April 2021]

CBS. 2020b. Minder Nederlanders voelen zich onveilig. Online. Available at: https://www.cbs.nl/nl-nl/nieuws/2020/10/minder-nederlanders-voelen-zich-onveilig [Accessed 29 April 2021]

CBS. 2020c. Minder traditionele criminaliteit, meer cybercrime. Online. Available at: Minder traditionele criminaliteit, meer cybercrime (cbs.nl) [Accessed 28 April 2021]

CBS. 2020d. Overheidsuitgaven en bestedingen; functies, transacties, overheidssectoren. Online, CBS StatLine. Available at: StatLine - Overheidsuitgaven en bestedingen; functies, transacties, overheidssectoren (cbs.nl) [Accessed 3 April 2021]

CBS. 2021a. Aanpassing Prijsindex Materiele Overheidsconsumptie (IMOC). Online, CBS Statline. Available at: https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/aanvullende-onderzoeksbeschrijvingen/aanpassing-imoc [Accessed 2 April 2021]

CBS. 2021b. Jaarmutatie consumentenprijsindex; vanaf 1963. Online. Available at: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70936ned/table?ts=1603819244855 [Accessed 2 April 2021]

CBS. 2021c. Geregistreerde criminaliteit; soort misdrijf, regio. Online. Available at : https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83648NED/table?fromstatweb [Accessed 24 April 2021]

CBS. 2021d. Geregistreerde criminaliteit; soort misdrijf, regio. Online. Available at: https://opendata.cbs.nl/#/CBS/nl/dataset/83648NED/table?searchKeywords=feitelijke%20aanranding%20der%20eerbaarheid [Accessed 24 April 2021]

CBS. 2021e. Overheidsfinanciën; kerncijfers. Online, CBS Statline. Available at: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/84114NED/table?ts=1615741491131 [Accessed 3 April 2021]

Chalfin, A. and McCrary, J. 2013. The effect of police on crime: New evidence from U.S. cities, 1960-2010. NBER Working Paper No. 18815.

Clemmensen, A. J. 2018. Quality adjustment of hospital services: Improving price and volume measures for health care. Statistics Denmark. Online. Available at: https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=SDD/CSSP/WPNA(2018)1anddocLanguage=En [Accessed 27 February 2021]

Cobbold, T. 2003. A Comparison of gross output and value-added methods of productivity estimation. Canberra: Productivity Commission Research Memorandum.

Council for the Judiciary. 2010. The judiciary system in the Netherlands. Online. Available at: https://www.rechtspraak.nl/SiteCollectionDocuments/The-Judiciary-System-in-the-Netherlands.pdf [Accessed 8 June 2021]

Council for the Judiciary. n.d. The Financing System of the Netherlands Judiciary. Online. Available at: https://www.rechtspraak.nl/SiteCollectionDocuments/The-Financing-System-of-the-Netherlands-Judiciary.pdf#search=financing%20of%20judiciary [Accessed 15 May 2021]

CPB. 2017. Toelichting bij kerngegevenstabel. Centraal planbureau. Online. Available at: https://www.cpb.nl/artikel/toelichting-bij-kerngegevenstabel [Accessed 6 May 2021]

CPB. 2018. Lange tijdreeksen overheidsfinancien. Online. Available at: https://www.cpb.nl/cijfer/lange-tijdreeksen-overheidsfinancien [Accessed 10 May 2021]

Cutler, D. M., Gosh, K., Messer, K., Raghunathan, T., Rosen, A. B. and Stewart, S. T. 2021. A satellite account for health in the United States. NBER Working Paper No. 27848.

De Jong, J. 2018. Het mysterie van de verdwenen criminaliteit. CBS. Online. Available at: 94582.PDF (politieacademie.nl) [Accessed 18 April 2021]

De Koning. 2017. Opheldering verzocht? Over de drastische daling van het aantal opgehelderde misdrijven. In A. G. Donker, R. A. Roks, B. Rovers, M.B. Schuilenburg, M. Smit, E. Snel (Eds.), De toekomstbestedingheid van de politie: 37-46. Den Haag: Boom Juridisch.

De Rechtspraak. 2006. Jaarverslag 2005: kwaliteit en kwantiteit. de Rechtspraak. Online. Available at: Jaarplannen, jaarverslagen en kengetallen | Rechtspraak [Accessed 10 April 2021]

De Rechtspraak. 2007. Jaarverslag 2006. Online. Available at: https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Raad-voor-de-rechtspraak/Jaardocumenten/Paginas/Archief-jaardocumenten.aspx [Accessed 10 April 2021]

De Rechtspraak. 2008. Jaarverslag 2007. Online. Available at: https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Raad-voor-de-rechtspraak/Jaardocumenten/Paginas/Archief-jaardocumenten.aspx [Accessed 10 April 2021]

De Rechtspraak. 2009. Jaarverslag 2008. Online. Available at: https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Raad-voor-de-rechtspraak/Jaardocumenten/Paginas/Archief-jaardocumenten.aspx [Accessed 10 April 2021]

De Rechtspraak. 2010. Jaarverslag 2009. Online. Available at: https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Raad-voor-de-rechtspraak/Jaardocumenten/Paginas/Archief-jaardocumenten.aspx [Accessed 10 April 2021]

De Rechtspraak. 2011. Jaarverslag 2010. Online. Available at: https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Raad-voor-de-rechtspraak/Jaardocumenten/Paginas/Archief-jaardocumenten.aspx [Accessed 10 April 2021]

De Rechtspraak. 2013. Jaarverslag 2012. Online. Available at: https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Raad-voor-de-rechtspraak/Jaardocumenten [Accessed 10 April 2021]

De Rechtspraak. 2014. Jaarverslag 2013. Online. Available at: https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Raad-voor-de-rechtspraak/Jaardocumenten [Accessed 10 April 2021]

De Rechtspraak. 2015. Jaarverslag 2014. Online. Available at: https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Raad-voor-de-rechtspraak/Jaardocumenten [Accessed 10 April 2021]

De Rechtspraak. 2016. Jaarverslag 2015. Online. Available at: https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Raad-voor-de-rechtspraak/Jaardocumenten [Accessed 10 April 2021]

De Rechtspraak. 2017. Jaarverslag 2016. Online. Available at: https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Raad-voor-de-rechtspraak/Jaardocumenten [Accessed 10 April 2021]

De Rechtspraak. 2018. Jaarverslag 2017. Online. Available at: https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Raad-voor-de-rechtspraak/Jaardocumenten [Accessed 10 April 2021]

De Rechtspraak. 2019. Jaarverslag 2018. Online. Available at: https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Raad-voor-de-rechtspraak/Jaardocumenten [Accessed 10 April 2021]

De Rechtspraak. 2020. Jaarverslag 2019. Online. Available at: https://jaarverslagrechtspraak.nl/ [Accessed 10 April 2021]

De Rechtspraak. n.d.a. Kwaliteit Rechtspraak verbeteren. Online. Available at: https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Raad-voor-de-rechtspraak/Kwaliteit-van-de-rechtspraak/Paginas/Kwaliteit-rechtspraak-verbeteren.aspx [Accessed 18 April 2021]

De Rechtspraak. n.d.b. Organisatie Rechtspraak. Online. Available at: https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie [Accessed 5 April 2021]

De Rechtspraak. n.d.c. Procedures Bestuursrecht. Online. Available at: https://www.rechtspraak.nl/Organisatie-en-contact/Rechtsgebieden/Bestuursrecht/Procedures [Accessed 8 June 2021]

Dean, E., Harper, M.J., and Sherwood, M.S. 1996. Productivity measurement with changing weight indexes of outputs and inputs. Paris: OECD Industry Productivity: International Comparison and Measurement Issues.

Diewert, W. E. 1976. Exact and superlative index numbers. Journal of Econometrics, 4(2): 115-145.

Diewert, W. E. 1992. The measurement of productivity. Bulletin of Economic Research, 44(3): 163–198.

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

Diewert, W. E. 2012. The measurement of productivity in the nonmarket sector. Journal of Productivity Analysis, 37(3): 217–229.

Diewert, W. E. 2017. Productivity measurement in the public sector: Theory and practice. Discussion Paper.

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

Drake, L. and Simper, R. 2003. An evaluation in the choice of inputs and outputs in the efficiency measurement of police forces. The Journal of Socio-Economics, 32(6): 701-710.

Dumaij, A., Niaounakis, T. and Urlings, T. 2014. Productiviteitstrends bij de rechterlijke macht- Onderzoek naar het effect van regulering op de

productiviteitsontwikkeling van rechtspraak en openbaar ministerie tussen 1965 en 2011. Delft: IPSE Studies.

Dunleavy P. and Carrera L. 2013. Growing the Productivity of Government Services. Cheltenham, UK: Edward Elgar.

Dunleavy, P. 2015. Public sector productivity: Puzzles, conundrums, dilemmas and their solutions. In J. Wanna, H. Lee and S. Yates (eds.), Managing under Austerity, Delivering under Pressure: Performance and Productivity in Public Services: 25-41. Australian Capital Territory: AnuPress [Australia and New Zealand School of Government (ANZSOG)].

Dunleavy, P. 2017. Public sector productivity: Measurement challenges, performance information and prospects for improvement. OECD Journal on Budgeting, 17(1): 153-179.

Eichhorn, W. and Voeller, J. 1976. Theory of the price index. Lecture Notes in Economics and Mathematical Systems 140. Berlin: Springer-Verlag.

Elbourne, A. and Grabska, K. 2016. Evidence on Dutch macroeconomic and sectoral productivity performance: Some stylized facts. CBP Netherlands Bureau for Economic Policy Analysis. Online. Available at: https://www.cpb.nl/sites/default/files/omnidownload/CPB-Background-Document-August2016-Evidence-on-dutch-macroeconomic-and-sectoral-productivity-performance.pdf [Accessed 9 June 2021]

European Commission for the Efficiency of Justice. 2016. Measuring the quality of Justice. As adopted on 7 December 2016, at the 28th plenary meeting of the CEPEJ. Online. Available at: https://rm.coe.int/1680747548#_Toc461720516 [Accessed 16 April 2021]

Eurostat. 2013. Glossary: Chain index. Online. Available at: Glossary:Chain index - Statistics Explained (europa.eu) [Accessed 13 April 2021]

Eurostat. 2014. Essential SNA: Building the system of national accounts. Online. Available at: https://ec.europa.eu/eurostat/statistics-explained/pdfscache/14452.pdf [Accessed 6 April 2021]

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

Eurostat. 2019. Manual on sources and methods for the compilation of COFOG statistics. Luxembourg: Publications Office of the European Union.

Fels, A. and Biggar, D. 2017. Competition in public markets. 16th Global Forum on Competition December 7th – 8th , Paris.

Fisher, 1. 1922. The making of index numbers. Boston: Houghton Mifflin.

Foxton, F., Grice, J., Heys, R. and Lewis, J. 2018. The welfare implications of public goods: Lessons from 10 years of Atkinson in the UK. The Sixth IMF Statistical Forum: Measuring Economic Welfare in the Digital Age: What and How? November 19th - 20th, Washington D.C.

Gemmell, N.; Nolan, P. and Scobie, G. 2017. Estimating quality-adjusted productivity in tertiary education: Methods and evidence for New Zealand. Working Paper 17/2017.

Genet, T. and Hayward, M. 2017. Productivity measurement study: Police. Research Note. New Zealand Productivity Commission.

Gielen, W. J. M. and Akkermans, M. M. P. 2019. Melding en aangifte. Online. Available at: Melding en aangifte | Veiligheid en slachtofferschap | Criminaliteit in Beeld [Accessed 3 May 2021]

Gilder, G. 1975. Public sector productivity. Public Productivity Review, 1(1): 4-8.

Government of the Netherlands. n.d. Organisation of the Dutch police. Online. Available at: Organisation of the Dutch police | Police | Government.nl [Accessed 15 April 2021]

Gollop, F.M. and Jorgenson, D. W. 1980. U.S. Productivity Growth by Industry, 1947-73. In J.W. Kendrick and B.N. Vaccara (eds.), New Developments in Productivity Measurement and Analysis: 17-136. Chicago: University of Chicago Press.

Hatry, H. 1978. The status of productivity measurement in the public sector. Public Administration Review, 38(1): 28-33.

Hauser, F. 2017. Quality of Public Administration: A Toolbox for Practitioners. Luxembourg: Publications Office of the European Union.

Heeks, M., Reed, S., Tafsiri, M. and Prince, S. 2018. The economic and social costs of crime: Second edition. Home Office. Online. Available at: The economic and social costs of crime (publishing.service.gov.uk) [Accessed 5 May 2021]

Helpman, E., Melitz, M. and Yeaple, S. 2004. Export versus FDI with heterogeneous firms. The American Review, 94(1): 300-316.

Home Office. (2001). Demonstration project. London: Home Office.

Hsieh, C. and Klenow, J. 2010. Development accounting. American Economic Journal: Macroeconomics, 2(1): 207-223.

Hulpverlening.nl. 2014. Nog altijd tekort aan wijkagenten. Online. Available at: https://www.hulpverlening.nl/nieuws/nog-altijd-tekort-aan-wijkagenten [Accessed 25 April 2021]

Hulten, C. H. 2000. Total factor productivity: A short biography. NBER Working Paper No. 7471.

Hulten, C. H. 2009. Growth accounting. NBER Working Paper No. 15341.

Inklaar, R., Timmer, M. and van Ark, B. 2008. Data for productivity measurement in market services: An international comparison. International Productivity Monitor, 16: 71-81.

Inspectie Gezondheidszorg en Jeugd. n.d. Wet bopz. Online. Available at: Wet Bopz | Wetten in ons toezicht | Inspectie Gezondheidszorg en Jeugd (igj.nl) [Accessed 15 April 2021]

IPSE Studies. n.d. Politie. Online. Available at : https://www.trendsinpubliekesector.nl/veiligheid-justitie/politie [Accessed 28 April 2021]

Jorgenson, D. W. 1995. Productivity : International comparisons of economic growth. Cambridge, Mass: The MIT Press.

Jorgenson, Dale W. and Fraumeni, B. 1989. The accumulation of human and nonhuman capital, 1948–84. In R. E. Lipsey and H. S. Tice (eds), The measurement of saving, investment and wealth: 227-282. Chicago: University of Chicago Press.

Jorgenson, D. W.; and Griliches, Z. 1967. The explanation of productivity change. The Review of Economic Studies, 34(3): 249–283.

Kim, H. Y. 1992. The translog production function and variable returns to scale. The Review of Economics and Statistics, 74(3): 546-552.

Kimbugwe, K., Lewis, R. and James, N. 2009. Labor inputs in public sector productivity: methods, issues and data. Economic and Labour Market Review, 3(4): 52-59.

Kohli, U. 2010. Labour productivity : average vs marginal. In W. E. Diewert, B.M. Balk, D. Fixler, K.J. Fox and A.O. Nakamura (Eds.). Price and productivity measurement: Volume 6 -- Index number theory: 103-132. Trafford Press.

Korf, D., Luijk, S. and de Meijer, M. 2018. Criminele samenwerkingsverbanden: Ontwikkelingen in aanpak en duiding effectiviteit. Amsterdam: Rozenberg Publishers.

Laspeyres, E. 187 1. Die berechnung einer mittleren waarenpreissteigerung. Jahrbücher für Nationalökonomie und Statistik, 16:296-314.

Jahrbiicherfur Nationulokononzie und Statistik, Vol. 16, pp. 296-314.

Lau, E., Lonti, Z. and Schultz, S. 2017. Challenges in the measurement of public sector productivity in OECD countries. International Productivity Monitor, 32: 179-194.

Lind, R. C. and Lipsky, J. P. 1971. The measurement of police output: conceptual issues and alternative approaches. Law and Contemporary Problems, 36(4): 566–588.

Lynch, R. 2006. Measuring non-market output in the national accounts. 29th General Conference of The International Association for Research in Income and Wealth.

Mandl, U., Dierx, A. and Ilzkovitz, F. 2008. The effectiveness and efficiency of public spending. European Commission. Online. Available at: the effectiveness and efficiency of public spending (europa.eu) [Accessed 16 May 2021]

McLellan, N. 2004. Measuring productivity using the index number approach: An introduction. New Zealand Treasury Working Paper, No. 04/05. Wellington: New Zealand Government, The Treasury.

Melitz, M. J. and Trefler, D. 2012. Gains from trade when firms matter. Journal of Economic Perspectives, 26(2): 91-118.

Moore, M. and Braga, A. 2003. The “bottom line” of policing: What citizens should value (and measure!). Online. Available at: Moore2003.indd (policeforum.org) [Accessed 30 April 2021]

Montesquieu, C. S., Nugent, T., and Alembert, J. L. R. 1899. The spirit of laws. New York: Colonial Press.

Nelson, R. R.; and Pack, H. 1999. The Asian miracle and modern growth theory. The Economic Journal, 109(457): 416-436.

Netherlands Council for the Judiciary. 2008. Quality of the judicial system in the Netherlands. Online. Available at: https://www.rechtspraak.nl/SiteCollectionDocuments/Quality-of-the-judicial-system-in-the-Netherlands.pdf [Accessed 10 April 2021]

New Zealand Productivity Commission. 2018. Measuring state sector productivity. Final report of the measuring and improving state sector productivity inquiry, vol. 2. Wellington: New Zealand Productivity Commission.

NOS. 2020. Rechtspraak loopt vast: meer dan 20.000 strafzaken uitgesteld. Online. Available at: https://nos.nl/artikel/2326024-rechtspraak-loopt-vast-meer-dan-20-000-strafzaken-uitgesteld [Accessed 15 April 2021]

OECD. 2001. Measuring productivity - OECD manual: Measurement of aggregate and industry-level productivity growth. Paris: OECD Publishing.

OECD. 2005. Superlative indices. Online. Available at: OECD Glossary of Statistical Terms - Superlative indices Definition [Accessed 8 May 2021]

OECD/Eurostat. 2012. Collective services in Eurostat-OECD methodological manual on purchasing power parities. Paris: OECD Publishing.

OECD. 2014a. Distinguishing between volume and price increases. In OECD, Understanding National Accounts: Second Edition: 47-77. Paris: OECD Publishing.

OECD. 2014b. Public sector. Online. Available at: https://stats.oecd.org/glossary/detail.asp?ID=2199#:~:text=The%20public%20sector%20comprises%20the,Context%3Aandtext=Moreover%2C%20the%20employer's%20identity%20is,on%20Public%20Sector%20Pay%20Trends. [Accessed 11 April 2021]

OECD. 2015. Government at a glance 2015. Paris: OECD Publishing.

OECD. 2019. Government at a glance 2019. Paris: OECD Publishing.

ONS. 2018. Quality adjustment of public service public order and safety output: current method. Office for National Statistics.

ONS. 2019. A guide to quality adjustment in public service productivity measures. Office for National Statistics.

Openbaar Ministerie. 2014. OM jaarbericht 2013. Online. Available at: https://www.om.nl/documenten/jaarverslagen/om/map/2012-t-m-2018/2013 [Accessed 3 April 2021]

Openbaar Ministerie. 2015. OM jaarbericht 2014. Online. Available at: https://www.om.nl/documenten/jaarverslagen/om/map/2012-t-m-2018/om-jaarbericht-2014 [Accessed 3 April 2021]

Openbaar Ministerie. 2016. OM jarbericht 2015. Online. Available at: https://www.om.nl/documenten/jaarverslagen/om/map/2012-t-m-2018/2015 [Accessed 3 April 2021]

Openbaar Ministerie. 2017. OM jaarbericht 2016. Online. Available at: https://www.om.nl/documenten/jaarverslagen/om/map/2012-t-m-2018/2016 [Accessed 3 April 2021]

Openbaar Ministerie. 2018. OM jaarbericht 2017. Online. Available at: https://www.om.nl/documenten/jaarverslagen/om/map/2012-t-m-2018/2017 [Accessed 3 April 2021]

Openbaar Ministerie. 2019. OM jaarbericht 2018. Online. Available at: https://www.om.nl/documenten/jaarverslagen/om/map/2012-t-m-2018/2018 [Accessed 3 April 2021]

Openbaar Ministerie. 2020a. Het OM in beeld. Online. Available at: https://www.om.nl/organisatie/openbaar-ministerie/documenten/brochures/om-brochures/om-in-beeld/2019/het-om-in-beeld [Accessed 1 April 2021]

Openbaar Ministerie. 2020b. OM jaarbericht 2019. Online. Available at: https://www.om.nl/documenten/jaarverslagen/om/map/2019-en-verder/om-jaarbericht-2019 [Accessed 3 April 2021]

Overheid.nl. n.d. Wet politiekosten evenementen. Online. Available at: https://www.internetconsultatie.nl/politiekostenevenementen/details [Accessed 27 April 2021]

Pierson, N. H. 1896. Further considerations on index-numbers. Economic Journal, 6: 127-131.

Politie. 2015. Jaarverslag nationale politie 2014. Online. Available at: Jaarverslag Nationale Politie 2014 | Jaarverslag | Rijksoverheid.nl [Accessed 20 April 2021]

Politie. 2016. Jaarverslag nationale politie 2015. Online. Available at: Jaarverslag Nationale Politie 2015 | Jaarverslag | Rijksoverheid.nl [Accessed 20 April 2021]

Politie. 2017. Jaarverantwording 2016. Online. Available at: Jaarverantwoording politie 2016 | Jaarverslag | Rijksoverheid.nl [Accessed 20 April 2021]

Politie. 2018. Jaarverantwording 2017. Online. Available at: Jaarverantwoording politie 2017 | Jaarverslag | Rijksoverheid.nl [Accessed 20 April 2021]

Politie. 2019. Jaarverantwording 2018. Online. Available at: Jaarverantwoording politie 2018 | Jaarverslag | Rijksoverheid.nl [Accessed 20 April 2021]

Politie. 2020. Jaarverantwording 2019. Online. Available at: Jaarverantwoording politie 2019 | Jaarverslag | Rijksoverheid.nl [Accessed 20 April 2021]

Politie. n.d.a. Over de Politie. Online. Available at : Over de politie | politie.nl [Accessed 10 April 2021]

Politie. n.d.b. Wat is de rol van de politie binnen de maatschappij?. Online. Available at: https://www.vraaghetdepolitie.nl/school/politiewerk/wat-is-de-rol-van-de-politie-binnen-de-maatschappij.html [Accessed 10 May 2021]

Pommer, E., and Eggink, E. 2010. Publieke dienstverlening in perspectief. SCP-memorandum voor de kabinetsformatie 2010. Den Haag: Sociaal en Cultureel Planbureau.

Pot, J. 2016. The Judiciary. In Ministry of the Interior and Kingdom Relations, The Dutch public service: 28. The Hague: Public Service Labour Affairs Directorate.

Pritchard, A. 2002. Measuring productivity change in the provision of public services. Economic Trends, 582: 20-32.

Rechters voor Rechters. n.d. Matters of principle: Codes on the independence and impartiality of justice. De Rechtspraak. Online. Available at: matters-of-principle.pdf (rechtspraak.nl) [Accessed 8 June 2021]

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

Schreyer, P. 2010. Towards measuring the volume of output of education and health services: A handbook. OECD Statistics Working Papers, No. 2010/02.Paris: OECD Publishing.

Schreyer, P. 2012. Output, outcome and quality-adjustment in measuring health and education services. Review of Income and Wealth, 58(2): 257-278.

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

Solow, R. M. 1957. Technical change and the aggregate production function. The Review of Economics and Statistics, 39(3): 312-320.

Somani, R. 2021. Public-sector productivity (part 1): Why is it important and how can we measure it?. World Bank. Online. Available at: https://documents1.worldbank.org/curated/en/913321612847439794/pdf/Public-Sector-Productivity-Part-One-Why-Is-It-Important-and-How-Can-We-Measure-It.pdf [Accessed 7 June 2021]

Statista. 2021. Euro (EUR) to british pound sterling (GBP) exchange rate from January 1999 to May 24, 2021. Online. Available at: https://www.statista.com/statistics/412806/euro-to-gbp-average-annual-exchange-rate/ [Accessed 28 April 2021]

Stroucken, M. 2016. The police force. In Ministry of the Interior and Kingdom Relations, The Dutch Public Service: 27. The Hague: Public Service Labour Affairs Directorate.

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

Terpstra, J. and Fyfe, N. 2019. Great expectations? Assessing the creation of national police organisations in Scotland and the Netherlands. International Journal of Police Science and Management, 21(2): 101-107.

The Commission of the European Communities. 2002. Commission decision of 17 December 2002. Official Journal of the European Communities, 347: 49-59.

The World Bank. 2020. Population, total – Netherlands. Online. Available at : https://data.worldbank.org/indicator/SP.POP.TOTL?locations=NL [Accessed 15 April 2021]

Thornhill, D. 2006. Productivity attainment in a diverse public sector. Public Seminar on Promoting Productivity in a Diverse Public Sector. Dublin: Ireland.

Tinbergen, J. 1942. Zur theorie der langfristigen wirtschaftsentwicklung. Weltwirtschaftliches Archiv, 55: 511-549.

Triplett, J. 2006. Handbook on hedonic indices and quality adjustments in price indices: Special application to information technology products. Paris: OECD Publishing.

Urlings, T. 2012. Productiviteitstrends bij de politie. Delf: TU Delft and IPSE Studies.

Urlings, T., Blank, J. and Niaounakis, T. 2014. Doelmatige dienders. Een vergelijking tussen regiokorpsen van de Nederlandse politie in de 21e eeuw. Delft: TU Delft and IPSE Studies.

Van der Ploeg, S. and de Wit, L. 2015. Ontwikkeling zaakzwaarte 2008-2014. Research Memorandum number 4/2015. Online. Available at: RM-2015-04-Ontwikkeling-zaakzwaarte.pdf (rechtspraak.nl) [Accessed 13 April 2021]

Vollaard. B. and Hamed, J. 2012. Why the police have an effect on violent crime after all: Evidence from the british crime survey. Journal of Law and Economics, 55: 901-924.

VVEM. n.d. Nieuwe wet politikosten mogelijk ook voor andere grote evenementen. Online. Available at: Vereniging van Evenementenmakers – Nieuwe wet politiekosten mogelijk ook voor andere grote evenementen (Bron: De Gelderlander) (vvem.nl)

Walsh, C. M. 1901. The Measurement of general exchange value. New York: Macmillan.

Yeong, S. 2019. The effect of police on crime and arrests: are police deterring or incapacitating criminals? Crime and Justice Bulletin, 223(223): 1–12.

Zouridis, S., Bouckaert, G.; van Roy, P., Stroobants, J., Crompvoets, V., Janssen, L. and Peeters, R. 2014. Politieproductiviteit - Triangulatie voor valide en betrouwbare productiviteitsmeting bij de politie. Tilburg University – KU Leuven.

Zouridis, S. 2015. Doe de goede dingen, maar doe ze wel steeds beter. het Tijdschrift voor de Politie, 77(7): 12-16.

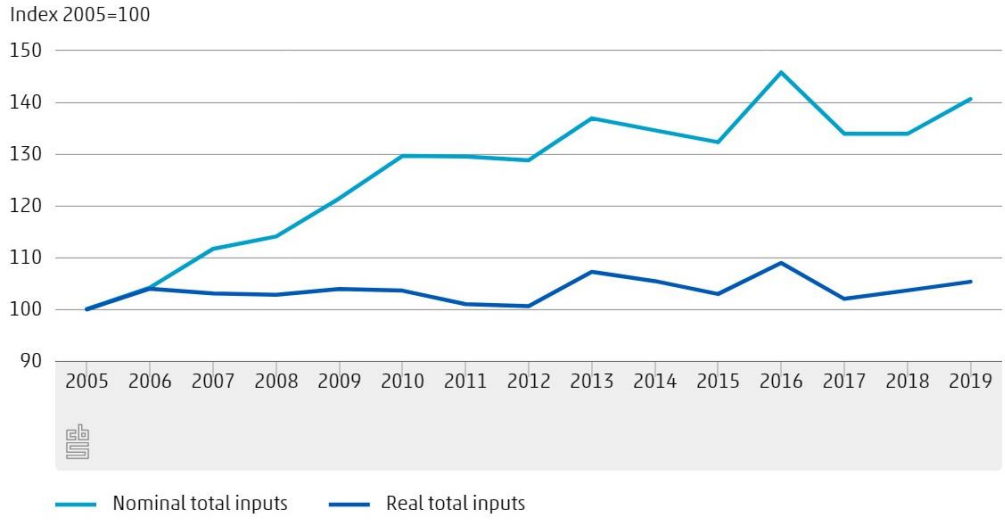
Appendices

A. Measuring labour inputs and input aggregation

Labour is crucial in most production processes. This is also the case for the three presented institutions. Employee remuneration constitutes the largest part of input costs. For all three institutions, direct labour measures are available which makes deflation unnecessary. According to Kimbugwe, Lewis & James (2009), using FTEs is a more preferable method to using a simple headcount. However, Kimbugwe, Lewis & James (2009), as well as Schreyer & Pilat (2001), point to the importance and benefits of quality-adjusted hours (actually) worked as a measure of labour volume. This measure more accurately reflects the productive services delivered by employees as a simple headcount neglects changes in average work times, for example, due to an increase in part-time work (Schreyer & Pilat, 2001). To properly investigate the question of productivity changes, additionally, the skill composition of labour has to be considered which indicates possible changes in the quality of labour. Generally, it is assumed that labour quality has increased over the past years and as such a quality-adjusted measure of labour input would increase faster than one that is not adjusted for quality and only takes into account working hours (Schreyer & Pilat, 2001).

While it is possible to distinguish between different employee groups in terms of FTE such as rechterlijke ambtenare ('judges and counsellors), directe and indirecte gerechtsambtenare ('direct and indirect civil servants') and landelijke diensten ('rural services') for the courts, wage information for the different labour categories is not available. Similarly, the public prosecution service and the police provide more granular information for different employee groups in terms of FTE but not in terms of wages. Therefore, it is not possible to adjust for changes in quality in terms of the composition of the workforce. More granular data concerning wage information for differently skilled employee groups would be desirable to account for possible quality changes in terms of labour and consequently, arrive at more accurate input figures. Once the different input costs are deflated if necessary the cost figures have to be aggregated. The assumption applied here in all three cases is that input factors are remunerated according to their marginal costs. Under this premise, the (deflated) inputs are aggregated according to their previous years cost shares in total input costs which are used as weights in the calculation. The practice of using cost shares as weights in aggregation follows from the growth accounting theory presented in the growth accounting equation (equation (2)). Note that the inputs are aggregated using a changing-weight index to take into account potential alterations in the relative input usage (Dean, Harper & Sherwood, 1996). Figure A.1 portrays the development of the inputs for the courts.

Figure A.1. Development of nominal and real inputs of the courts



It is visible that real inputs overall increased between 2005 and 2019. Nevertheless, after the maximum value in 2016, inputs decreased again.

The development of the inputs for the public prosecution service (figure A.2) looks quite different. Inputs (nominal as well as deflated) first decreased from 2013 to 2014. Afterwards, both input series saw an increase but while real inputs exceed their 2013 level in 2014, nominal inputs do not. Overall, real inputs increased by 12 percent from 2013 to 2019.

Figure A.2. Development of nominal and real inputs of the public prosecution service

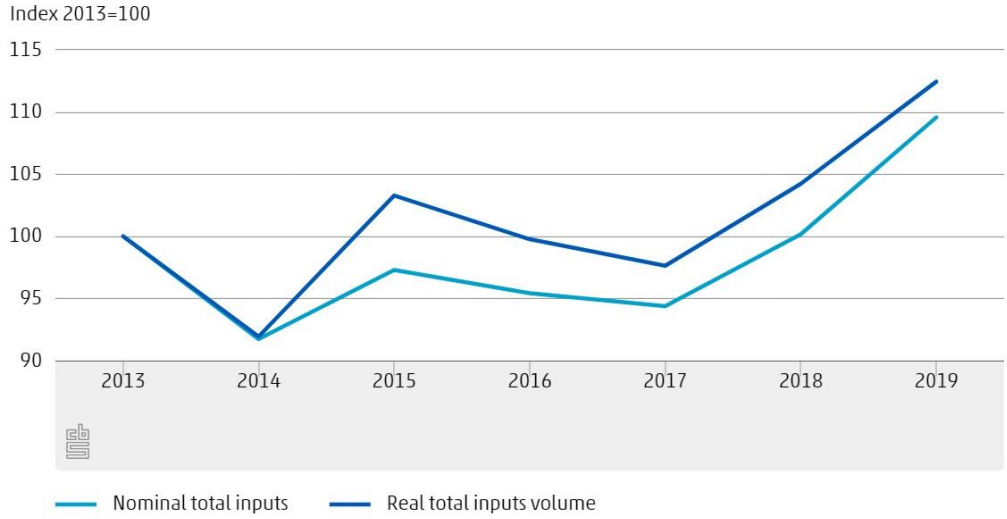
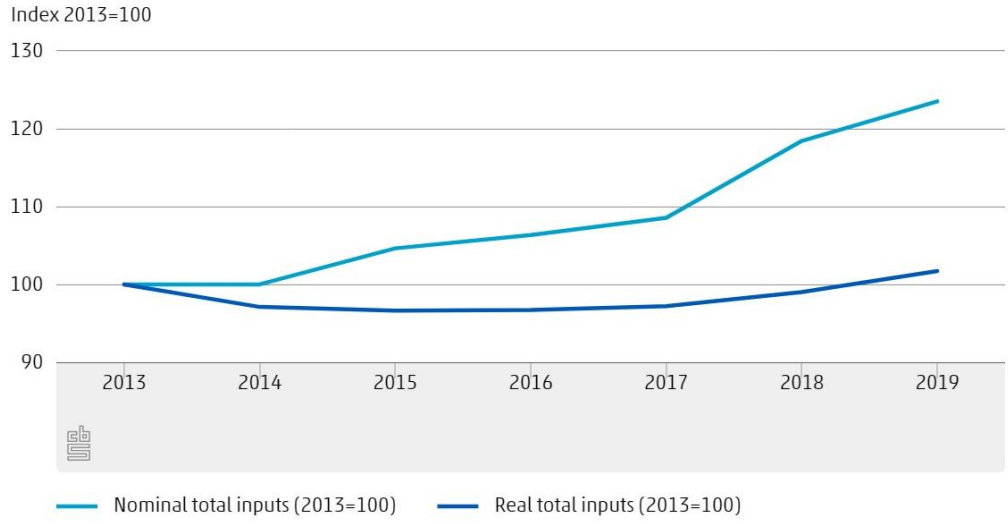


Figure A.3. Evolution of nominal and real inputs of the police



Lastly, Figure A.3 portrays the inputs of the police sector. It seems as if real inputs have not changed much over the examined period. However, this illustration can be misleading. Table A.4 below shows that the input factors capital expenditure and current expenditures have seen large changes over time. Nevertheless, their share in total input costs is relatively small. Labour, which makes up around 70 percent of total input costs, has seen almost no change between 2013 and 2019. Hence, the evolution of input costs as visible in figure A.3 presents only small changes. In combination with the mostly decreasing outputs, the almost constant inputs result in decreasing TFP for many different weighting options and scenarios as presented in section 4.3.4.

Figure A4: Change in different input factors of the police, 2013-2019

	Intermediate consumption	Labour	Other capital expenditures	Other current expenditures	Cost of volunteers
%	-4.73	-0.50	34.59	-40.98	-9.51

B. Background on index numbers

As it is not possible to simply sum inputs and outputs to arrive at aggregate measures, statisticians often utilize index number theory in productivity measurement. To provide an example of productivity measurement, take a firm that produces many outputs with quantity q_i and price w_i by using different inputs where p_i denotes the price of inputs and x_i denotes the quantity of inputs. To estimate productivity growth, it is necessary to observe each component over time, i.e. in at least two periods, the base period and period t (Simpson, 2009). Then one

can obtain the Laspeyres input quantity index Q_I and the respective output quantity index Q_O as follows[33]:

$$Q_I = \sum_{i=1}^{N} p_i^1 x_i^t \Big/ \sum_{i=1}^{N} p_i^1 x_i^1 \tag{21}$$

$$Q_o = \sum_{i=1}^{M} w_i^1 q_i^t \Big/ \sum_{i=1}^{M} w_i^1 q_i^1 \tag{22}$$

Consequently, TFP can be calculated by dividing the quantity index of output by the quantity index of inputs[34]:

$$TFP = Q_O / Q_I \tag{23}$$

Note that all of these indices can be either fixed base or chained based. In the case of fixed based indices, all values are related to the same fixed base period while for a chained base index the value is related to the previous year (Eurostat, 2013)[35].

C. Axiomatic tests for indices (based on Diewert, 1992)

Table C.1 presents the four most commonly used axiomatic tests and the indices that pass each test (Diewert, 1992). It is to be highlighted that this table presents price indices but the same tests can be conducted for quantity indices (see McLellan, 2004)

Table C.1. Axiomatic Test for most common indices (based on Diewert, 1992)

Test	Mathematical Test	Price indices passing the Test
Identity or Constant Prices Test (Laspeyres, 1871)	The price index should be equal to 1 if all prices are identical in the two periods, regardless of what the quantities are in the two periods	Paasche, Laspeyres, Fisher, Törnqvist-Theil
Constant Quantities Test (among others Walsh, 1901)	If quantities remain stable over the two examined periods, then the ratio of price indices between the two periods should approximate the ratio of values between the two periods	Paasche, Laspeyres, Fisher

33 By taking the current price instead of the quantity, one can calculate the Laspeyres input and output price indices.
34 Similar calculations can be conducted using the other 3 indices.
35 Following this, for the first two periods the fixed based and chain based indices are the same and discrepancies only arise from period 3 onwards (Diewert, 2017).

Table C.1. Continue

Test	Mathematical Test	Price indices passing the Test
Proportionality in Period t prices Test (Walsh, 1901; Eichhorn and Voeller, 1976)	If all prices in period t are multiplied by a constant larger than 0, the price level in period t relative to the price level in period s rises by the same constant larger than 0	Paasche, Laspeyres, Fisher, Törnqvist-Theil
Time Reversal Test (Pierson, 1896; Walsh, 1901; Fisher, 1922)	If prices and quantities for the two periods are interchanged, the price index should be the reciprocal of the original price index	Fisher, Törnqvist-Theil

D. Deflators

To use nominal cost figures in the productivity calculations deflation is necessary. Hereby, the choice of suitable deflators is paramount (Atkinson, 2005). Ideally, intermediate consumption should be deflated in a product-by-product manner (Eurostat, 2014). However, this exceeds the scope of this analysis and the scope of the available data. Therefore, this analysis will utilize the *Index Materiële Overheidsconsumptie* ('price index for the material consumption of the government'), also called the *IMOC.* This index is available on *StatLine* (CBS, 2021e) and is also used by the *Centraal Planbureau* and the *Ministry of Finance* (CBS, 2021a).

Government investments will be deflated using the *Index Bruto Overheidsinvestering* (*IBOI*) ('price index for gross government investments'). This index provides information about the average price increases in investments conducted by the collective sector, among which the government falls (CPB, 2017). Similar to the IMOC, the IBOI is available on *StatLine* (CBS, 2020a).
For taxes and other current expenses, the *Consumer Price Index (CPI)* can be used as a deflator (CBS, 2021b).

Even though the analyses tried to apply direct labour measures as much as possible, these are not always available. For example, externally hired labour or volunteers costs require the application of a suitable deflator. The deflators for government wages can be calculated from the sector accounts available on StatLine (CBS, 2016; 2019b). However, the sector accounts are only available in current prices. Therefore, to exclude the changes in wages caused by inflation, the real sector accounts wages are derived using a deflator calculated from the input-output tables (CBS, 2016; 2019b). These are available in current and past year prices. By deflating the sector accounts wages, it is possible to calculate a fitting deflator. This shows the changes in labour expenses over time. However, the figures are only available at the aggregated government sector-level and therefore, do not account for changes in wages of the different institutions which are part of this analysis.

E. Overview social costs of crime

Table E.1. Overview of social costs of crime and categorisation

Categories in Heeks et al. (2018)	Social costs as consequences of crime in 2015		Crime category in author's categorisation	Average costs per group used for calculations	
	British £	Euro[36]		Euro (2015)	
homicide*	2 343 730	3 210 589	3 – Violent Crime	17 685	3 210 589
violence with injury	11 220	15 370	3 – Violent Crime	17 685	
violence without injury	3 760	5 151	3 – Violent Crime	17 685	
rape	31 440	43 068	3 – Violent Crime	17 685	
other sexual offences	5 220	7 151	3 – Violent Crime	17 685	
robbery	6 310	8 644	1 – Property Crime	3 817	
domestic burglary	3 410	4 671	1 – Property Crime	3 817	
theft of vehicle	4 660	6 384	1 – Property Crime	3 817	
theft from vehicle	590	808	1 – Property Crime	3 817	
theft from person	920	1 260	1 – Property Crime	3 817	
criminal damage - arson	3 110	4 260	2 - Vandalism	2 658	
criminal damage - other	770	1055	2 - Vandalism	2 658	
fraud	830	1137	1 – Property Crime	3817	

* The category of homicides receive two values in the last column because when calculating the weights for registered crimes, homicides are calculated separately and thus the real value and not average value for the group is used (see equation 19).

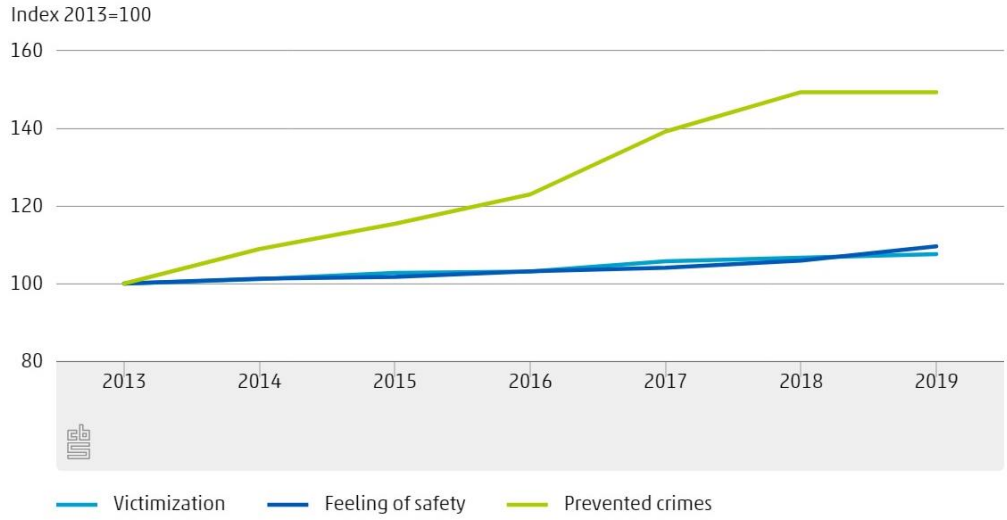
F. Changes in societal outcome of safety

Figure F.1 portrays the changes in certain societal outcomes that are related to the degree of safety. The index of prevented crimes from the Trends in Publieke Sector (TiPS) initiative is an approximation of the degree of crimes prevented. The index is calculated by dividing the expected crimes by the actual crimes, where the number of expected crimes is derived by using socio-economic indicators such as unemployment rates and demographics (IPSE Studies, n.d.). Further, the degree of safety is influenced by the level of crime. It might seem straightforward to deduct the level of safety from the amount of registered crime. Nevertheless, it should be considered that much of the actual crime happening is not reported to the police. In 2017, only a third of the crime experienced by citizens was reported (Gielen & Akkermans, 2019). Therefore, the level of victimization is a better indicator of the safety level than the number of registered crimes. The number of people experiencing traditional crimes such as property crimes, violent crimes and vandalism is reported by CBS (2020c). Naturally, the inverse of the index is taken and the blue line in Figure F1 shows the index of how many people have not fallen victim to a crime so that a positive trend also shows a positive development in terms of the desired outcome. Lastly, Figure F.1 portrays the feeling of safety of Dutch citizens. This data can be obtained from the CBS (2020b). Naturally, the

36 The utilised average annual exchange rate in 2015 was 0.73£/€ (Statista, 2021).

results from surveys are subjective. Two citizens living in the same neighbourhood can have very different perceptions of safety.

Figure F.1. Changes in outcomes in terms of safety



All outcome indicators in Figure F.1 exhibit a positive trend, indicating that according to the utilized indicators, the level of safety is increasing in The Netherlands. However, since it is not possible to know which part of the changes in outcomes can be ascribed to the work of the police, as other factors such as private crime prevention also contribute, such outcome indicators, for now, cannot be incorporated in analyses of police productivity.

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