

An appraisal of Public Oversight Body (POB) inspections of statutory audit firms in the EU: the case of the Netherlands

Abstract

This paper finds that the quality of statutory audit firm inspections by Public Oversight Bodies (POBs) in the EU is problematic. These POBs determine the quality of statutory audit firms as follows: (1) they inspect an audit firm's quality control system, in part based on (2) the inspection of a sample of individual statutory audits of these audit firms. They then combine the (1) and (2) outcomes to arrive at an overall judgement of the quality control system of an audit firm. In this paper the focus is on (2): POB inspections in the EU of a sample of individual statutory audits.

The paper further focuses on the case of the Netherlands, one of the EU countries with a distinctly critical POB. The Dutch POB's (AFM) findings and criticism since 2010 have led to a still continuing discussion about necessary reforms of the Dutch audit sector.

The empirical part of this paper is a case-study. Case study empirical evidence is used to critique the AFM oversight approach. The case study finds that the approach the AFM uses for selecting and testing of individual audit files to help evaluate an audit firm's quality control system is flawed. It is not 'sound theory and evidence' based. Since all EU POBs use a version of the same approach, the critique extends to those POBs as well. In fact, worldwide, audit firm POBs are members of IFIAR (International Forum of Independent Audit Regulators). IFIAR advocates approaches to audit files inspections such as that used by the AFM.

Recommendations for a more rational POBs audit files inspection approach are sketched.

Finally, the paper discusses why the AFM is not operating in a way supported by theory and evidence.

1. Introduction

This paper finds that the quality of statutory audit firm inspections by Public Oversight Bodies (POBs) in the EU is problematic. The POBs determine the quality of these audit firms as follows: (1) they inspect an audit firm's quality control system, in part based on (2) the inspection of a sample of individual statutory audits of these audit firms. The POBs then combine the (1) and (2) outcomes to arrive at an overall judgement of the quality control system of an audit firm. The statutory audit firms are POB registered audit firms that have passed a required initial screening by the POB.

This approach follows the EU 2006 Auditing Directive. That Directive replaced, in most EU member states, peer firm quality reviews of audit firms by independent POB oversight.

In a number of EU countries, most notably the UK and The Netherlands, the POB has since voiced strong criticism of the audit firms they oversee.

It is therefore important to ask: how well do POBs do (1) and (2)? In this paper the focus is on (2): the POB inspection of a sample of individual statutory audits. It further focuses on the case of the Netherlands, one of the EU countries with a distinctly critical POB. The Dutch POB's criticism has led to a still continuing discussion about necessary reforms of the Dutch audit sector.

The advent of independent oversight of the audit sector in the Netherlands by a POB, the AFM ('Autoriteit Financiële Markten'), is a consequence of the Dutch implementation of the 2006 EU Auditing Directive and Regulation. The Directive and Regulation were a revision of the earlier EU 8th 1984 Company Law Auditing Directive. That 2006 revision grew out of concerns with audit quality in the wake of the Enron (in the US) and the Parmalat (in the EU) accounting scandals. A subsequent revision has led to the current 2014 EU Auditing Directive and the 2014 Audit Regulation. The 2014 Auditing Directive, and the linked Regulation, is now implemented in all 27 EU member states and the UK, as well as in Iceland and Norway. See Accountancy Europe, (2018) for a survey: for alternative surveys see Offermans, Vanstraelen (2014); Osma et al (2017), Carson et al (2021) and Florou, Shuai (2021)).

In the Netherlands the relevant legislation, since 2006, is the 'Wet Toezicht Accountantsorganisaties' (WTA) and the 'Besluit Toezicht Accountantsorganisaties' (BTA). WTA (2006) and BTA (2006) were in part based on a draft of the 2006 EU Auditing Directive and Regulation. In 2008 the WTA and BTA were brought into line with the Directive (Pouw, Maijoor (2008)). WTA (2008) provides the context for this paper. It was again updated, to WTA (2016) based on the 2014 Auditing Directive and Regulation. BTA (2016) is also based on the 2014 Directive and Regulation. Both WTA and BTA were updated again in 2018. This reflects the continuing discussion about audit quality and regulation in the Netherlands. A further update is pending.

The EU 2014 (and 2006) Auditing Directive instructs a member state POB (article 29 and 32 f) to carry out audit firm inspections to determine the quality of the audits performed. That is: review the audit firm's quality control system, *supported* by an inspection of individual audits carried out by the audit firm. Dutch law (WTA (2008), and the revisions since), formulates the required setup of an AFM review slightly differently (Article 48a): review the audit firm's quality control system, *at least* by inspecting individual audits carried out by the audit firm. Note the difference between '*supported by*' and of '*at least by*'. This difference plays a role in the discussion below.

This paper finds that the approach the AFM uses for selecting and testing of individual audit files to help evaluate an audit firm's quality control system is flawed. Since all EU POBs use a version of the same approach, the critique extends to those POBs as well. In fact, worldwide, audit firm POBs are members of IFIAR. IFIAR advocates approaches to audit file inspections such as that used by the AFM.

Recommendations for a more rational POBs audit files inspection approach are sketched.

The empirical part of this paper is a case-study. Case study empirical evidence is used to critique the AFM oversight approach. Regulatory oversight should be 'rational', that is 'sound theory based' and 'evidence-informed' (see Buijink (2006), Leuz (2019)). The current AFM's oversight of Dutch audit firms is not rational in that sense. It appears that there are insufficient incentives in place that would cause AFM's statutory audit firm oversight to be rational.

2. Public oversight of audit firms in the Netherlands

2.1. Preliminary remarks: institutional details

(a) The AFM (e.g. in AFM (2018)) uses the term 'audit sector' to denote the 'audit industry': i.e. client firms demanding audits and audit firms supplying these. This terminology will be used throughout this paper.

(b) The EU distinguishes corporations that have a public interest character. They are called Public Interest Entities (PIE). These are listed corporations and corporations in the financial industry: banks and insurers. In Dutch they are called 'Organisaties van Openbaar Belang (OOB). Hence, Dutch domiciled Amsterdam Stock Exchange listed corporations are PIEs. Note that, and this varies by EU member state, certain non-corporate entities (e.g. certain non-profits) can also be a PIE. For the auditing of the PIEs the oversight regime is stricter.

(c) In the EU a distinction needs to be made by statutory financial statements audits, and voluntary audits. In most EU member states, as is the case in the Netherlands, audits are mandated (i.e. statutory) for medium-size and large corporations and certain categories of other organisations. In a number of EU countries they are also mandated for small corporations. The POBs in the EU focus their oversight on statutory audits by registered (by the POB) statutory audit firms. These audit firms can of course also carry out voluntary audits.

(d) Note that the US external audit sector POB, the PCAOB, was created in the US in 2003, post Enron, by the 2002 Sarbanes Oxley Act. The EU audit sector POB's (including the one in the UK) are clearly modelled on the PCAOB, certainly in their inspections approach. One important difference is that in the US the PCAOB only inspects audit firms that audit listed, SEC registered, corporations. Comparing the Netherlands and the US, focusing on the Big 4 audit firms, the numbers are (for the year 2014: why 2014?: on that see below) are as follows. The number of (statutory) audits that year by the Big 4 in the Netherlands was 11.248 (AFM (2014)), in the US it was 3.061 (Audit Analytics (2014)). Hence: counterintuitively perhaps, the AFM oversees a larger number of audits of the Big 4 audit firms than does the PCAOB.

(e) Currently (2021), the Dutch audit sector looks as follows. The NBA ('Nederlands Beroepsorganisatie van Accountants') is the sole certified auditors' association. Total NBA membership is 22K, of which 9K work in statutory auditing practice and about 1700 of them are AFM registered engagement partners (mostly equity and salary partners) that can do statutory audits. There are 272 AFM registered audit firms that can do statutory audits (though not all of them do). There are 6 audit firms that can do PIE audits (among them the Big 4). Auditees, clients, are 17K corporations and 2.5K nonprofit entities: hence, there are about 19.5 K statutory audits in total. Almost 900 of these entities are PIEs. The AFM has delegated non-PIE statutory audits audit firms oversight to the NBA and the SRA. The SRA ('Samenwerkende Registeraccountants en Accountants Administratie consulenten') has a membership of 191 non-PIE AFM registered audit firms. The NBA oversees the remaining 81 audit firms. This arrangement will end in 2022, after which the AFM will oversee all statutory audits and all PIE and non-PIE audit firms.

The NBA estimates (NBA (2019), p7) that additionally there are 1700 other, non-AFM registered, audit firms in which NBA members are involved, and that can do voluntary audits. These audit firms are potential entrants into the Dutch statutory audit sector.

(f) Note that in the EU there also exist, also mandated by the 2006 Auditing Directive, internal statutory audit inspections: within the audit firm itself. These come in two forms: Engagement Quality Reviews (EQR) that are concurrent engagement quality reviews (mandated for PIE audits) and Internal (Engagement) Quality Reviews (IQR) of completed audits (mandated for PIE audit firms). In Dutch these internal inspections are called OKB ('Opdrachtgerichte Kwaliteit Beoordeling') and IKO ('Intern Kwaliteit Onderzoek'). The English acronyms will be used in this paper. Obviously, the IQRs are similar in intent to the external AFM inspections, but they also may have a Human Resources aspect. While this paper focuses on the AFM inspection, its findings are also relevant for IQRs. Note however that IQRs tend to focus on engagement partners (and play a role in internal audit firm Human Resources policy). POB inspections focus on audit firms. How IQRs are set up in PIE audit firms can be observed in the Transparency Reports that these firms publish annually: also as a consequence of the 2006 Auditing Directive. EQRs are different in character. Any issue between the engagement auditor and the EQR auditor will have to be resolved before the audit opinion is communicated to the client. As explained, EQRs are mandated for audits of PIEs, and where relevant (to be determined by the audit firm), also for non-PIE client statutory audits.

(g) The Netherlands has, since 2009, an independent auditor court.

(h) Auditors and audit firms in the Netherlands use the IAASB auditing standards since the Auditing Directive and WTA in 2006.

(i) Finally: since 2006 the Dutch Ministry of Finance has been the designer and preparer of legislation for the audit sector.

2.2. Auditing Directive related AFM external inspections

The question that is addressed in this paper is, how, and how well, does the AFM in the Netherlands carry out (2): the testing of individual audit files to (in part) help determine the quality of a Dutch audit firm's quality control system.

Before 1996 there were no external inspections of audit firms and the audits they carried out in the Netherlands. From 1997 audits carried out in the Netherlands were subjected to peer review organised by the professional bodies for accountants in the Netherlands: NIVRA and NOVAA (that merged in 2013 to form NBA). Teams from other Dutch audit firms performed these. Since 2007 the external quality inspections have been carried out by the AFM (van Oppijnen (2012)).

The AFM carries out its audit firms, hence audit sector, oversight as follows. First, as explained, it continuously vets and registers the audit firms (both PIE and non-PIE audit firms) active in the sector. PIE audit firm registration is stricter. The AFM also registers all statutory auditors that can carry out statutory audits: these are the 1700 engagement partners mentioned earlier. Second, and this is central to this paper, the AFM is tasked with inspecting, at least once every three years, the quality of the

statutory audits carried out by the audit firms that perform PIE statutory audits. The AFM actually does this with three year intervals. The other registered, non-PIE, statutory audit firms are to be inspected, as explained earlier, at least once every six years. The NBA and the SRA, in every year of that six year cycle, inspect a subset of the non-PIE AFM registered audit firms. This creates an unpredictability for the inspections of the non-PIE audit firms that does not exist for the PIE audit firms.

In the 2006 and 2014 Auditing Directive article 29 (intro and paragraph f) says: "Each member state shall ensure that all statutory [...] audit firms are subject to a system of [POB] quality assurance [...] the scope [of which] includes an assessment of [...] the[ir] internal quality control system [...], *supported [Italics added]* by adequate testing of selected audit files". Paragraph h of article 29 stipulates that the "quality assurance reviews shall take place on the basis of [...] risk".

In WTA (2008 and 2016) this is formulated as follows in article 48a, 1. to 3. : "1. The AFM shall assess at least once every six years, or as many times as needed based on a risk analysis, whether an audit firm complies with the provisions under or pursuant to this Act and the EU Regulation. 2. If an audit firm conducts statutory audits with public-interest entities, the AFM shall assess, in derogation from the first paragraph, at least once every three years whether the entity complies with the provisions laid down in or pursuant to this Act. 3. The AFM shall base its assessment *at least [Italics added]* on an assessment of a selection of audit files.". This article 48a first appeared in WTA (2008).

Note the difference between 'supported' (Directive) and 'at least' (WTA 2008).

With regard to the selection of audit files to be inspected, as required in both the Audit Directive and the WTA the European Commission issued a clarifying Recommendation in 2008 (European Commission (2008)). With regard to testing of selected audit files: "..., at least a significant part of audit files should be selected on the basis of an analysis of the risk of an inadequate execution of the statutory audit.". That is, the European Commission requires the inspection of primarily the 'riskier' audit files.

In addition WTA (2008) contains a statutory audit firm 'duty of care' article (WTA (2008), article 14), which is not in the 2006 Auditing Directive. Statutory audit firms have a duty of care towards stakeholders, such that their engagement partners carry out the audits with sufficient quality. This in effect means that the AFM inspects both an audit firm's 'duty of care performance' and the audit firm's 'internal quality control system performance' (article 18 WTA). This is an odd arrangement. Various authors. e.g. Wallage et al (2020), and Eijkelenboom, van Opijnen (2020) have also noted this. It will play an important role below.

Note that the AFM, since AFM (2014), publishes the individual PIE audit firm inspections findings. It does this immediately. There is no publication delay to enable a 'remediation'.

4. The AFM audit quality inspections and reports

4.1. PIE statutory audit firm inspections

The AFM has published seven PIE statutory audit firm inspection reports to date: AFM (2010): B4 audit firms; AFM (2013a): all 9 non-B4 PIE audit firms; AFM (2014): B4 audit firms; AFM (2015): all 9 PIE audit firms; AFM (2017): B4 audit firms, and AFM (2019): all 5 non-B4 PIE audit firms. Finally: AFM (2021): all 6 PIE audit firms. Note that in AFM (2021) the AFM changed its inspection sampling approach. It now selects audit files to be inspected from a list of IQRs carried out by the 6 PIE audit firms.

Note that the audit files inspected are from 1 or 2 years before the publication year of the report. The AFM inspects completed audit files.

These are the results for the files inspections:

- 1) AFM (2010): B4 audit firms: 46 files inspected: 63% defective (quality control systems and files inspections) [*around 11000 B4 audits in total*]
- 2) AFM (2013a): all 9 non-B4 PIE audit firms: 47 files inspected: 74% defective (quality control systems and files inspections) [*around 3000 audits in total*]
- 3) AFM (2014): B4 audit firms: 40 files inspected: 45% defective (only file inspections) [*around 11000 audits in total*]
- 4) AFM (2015): all 9 PIE audit firms (only quality control systems)
- 5) AFM (2017): B4 audit firms: 32 files inspected: 60% defective (only files inspections) [*around 9000 audits in total*]

6) AFM (2019): all 5 non-B4 PIE audit firms: 14 files inspected: 80% defective (quality control systems and files inspections) *[around 3000 audits in total]*

7) AFM (2021): 6 PIE audit firms: 18 files inspected: 33% defective (only files inspections) *[around 9000 audits in total]*

Note the decoupling of the quality control systems inspections and audit files inspections. The AFM does not explain this in its inspection reports, but the decoupling appears to be based on the difference in the Auditing Directive and the WTA mentioned above. As pointed out above the Auditing Directives requires statutory audit firm quality control system inspections *supported by* files inspections. WTA requires statutory audit firm quality control system inspections *at least based on* files inspections.

Also note the small number of inspected audit files relative to the number of audits carried out by the audit firms inspected.

Implicit in the AFM reports involving individual files inspections is that none of the Dutch statutory audit PIE firms so far has scored 'satisfactory' as an AFM inspection outcome. This contrasts with the outcome for the Dutch statutory non-PIE audit firms described in the next section (4.2).

In all AFM reports with files inspections the AFM explicitly states that finding defects in an audit file does not invalidate the financial statements of the client involved.

Moreover, while the AFM can bring the engagement partners involved in defective audits to the Dutch* Auditor Court, to date it has done this only once (in 2022).

From the start in 2010, all AFM reports show a high 'defective audits' percentage. This 'defective audits' percentage has caught the attention of observers of the audit sector since 2010.

AFM (2014) in particular created substantial financial press coverage. It also drew parliamentary attention and parliamentary debate. As a consequence the NBA, also in 2014, issued a detailed report outlining 53 measures meant to improve statutory audit quality. The NBA also created a Monitoring Committee Accountancy (MCA) that subsequently issued 3 reports highly critical of audit quality in the Netherlands.

AFM inspections reports outcomes after 2014 provided additional input for criticism by the MCA. This continuing debate led to the Ministry of Finance to create yet another advisory committee, the 'Commissie Toekomst Accountancysector' (CTA) in 2019. The CTA final report, CTA (2020), was used by the Ministry of Finance to generate a list of 20 further audit sector reforms and to introduce to Parliament a proposal for a 'Wet Toekomst Accountancy'. This is legislation that is intended to again amend the WTA and BTA. AFM (2014) also led to the B4 statutory audit firms to be fined two years later: in 2016.

Hence the AFM inspection reports have substantially impacted the 'regulation of the audit sector' debate in the Netherlands and has led to a considerable number of

suggestions of additional regulatory actions. It is therefore important to investigate the quality of the AFM inspections.

This paper uses a case-study to do this. AFM (2014) led in 2016 to AFM fines for all B4 audit firms involved. Two of the B4 appealed to the fines. This paper uses one of these appeals (in 2 instances) as its case-study material.

Note the following: there is a second selection the AFM makes within each inspected audit file as to the topics in the file that are inspected. That is, the AFM also samples topics within a client file.

Finally: the AFM can also use auditing 'incidents' it learns about, in various ways, as a basis for an inspection. It has only done so in AFM (2010).

4.2. Non-PIE audit firms inspections

The AFM has delegated, covenant based in both cases, non-PIE audit firm inspections of statutory audits to committees within the SRA and the NBA. The SRA is a trade organisation of medium sized and small audit firms (191 of its member audit firms are currently registered with the AFM). The remaining none-PIE audit firms registered with the AFM are inspected by the NBA committee. In 2020 (most recent year available) the SRA and NBA committees inspected 55 audit firms: out of 246 non-PIE active statutory audit firms. Inspections combine an inspection of the quality control system in place and an inspection of selected audit files. The overall outcome at the audit firms level is 'satisfactory' or 'not-satisfactory'.

In 2020 the number of 'not-satisfactory' scoring audit firms was 23. Such a score triggers a re-inspection in the next year, and may eventually result in the audit firm in question leaving the statutory audit sector, i.e. an AFM deregistration.

The method for the selection of audit files within the audit firms follows the method the AFM uses for the selection of audit files in the PIE audit firms. However, in this case a selection of the non-PIE audit firms to be inspected is also necessary. It also appears to be higher-risk oriented. This creates a further sampling issue.

The individual audit firm files inspection outcomes were not published in the NBA and SRA inspection reports.

From 2022 the arrangement with the NBA and SRA will end, and the AFM will also oversee the non-PIE statutory audit firms.

5. The quality of AFM audit firm inspections: a case study

5.1. Introduction

This paper critiques the AFM's audit files inspection methodology with reference to the AFM inspection of the 2012 statutory audits of the B4 audit firm EY. The inspection result was published in AFM (2014). On the basis of the outcome the AFM fined EY in 2016. Its motivation is in AFM (2016). The fine was based on the outcome of files inspections only (as explained above; see also Eijkelenboom, van Oppijnen (2020), and explicitly also the AFM itself in AFM (2016, p4)). That is, the

AFM, in AFM (2014), did not inspect EY's system of quality control. It used EY's 'duty of care' obligation to 'go' from the files inspection outcome directly to a negative EY 'system of quality control quality' outcome. The fine was EURO 2.230.000.

As mentioned, all B4 audit firms were fined. KPMG and Deloitte accepted the fine. PwC and EY appealed and won in a court of first instance in 2017. On appeal by the AFM, PwC and EY won their case again in 2019. The fines were therefore rescinded. In the course of these two cases more detailed information became available about the methodology that the AFM uses.

The focus here is on the EY case: Vonnis (2017) and Vonnis (2019). In the EY case the AFM files inspections method was discussed in more detail because EY sought statistical advice. It therefore provides more informative case study material. That material is in AFM (2016), which explains the decision to fine EY, and in Vonnis (2017) and Vonnis (2019).

EY, as did PwC, won both of its appeals because in both instances the courts objected to the AFM not having also inspected the audit firm's system of quality control. It had based its fine on the outcome of the audit files inspection only. In other words both courts followed the Audit Directive 'logic'.

However the courts did not explain this explicitly. Nor did they mention specifically the differences noted above between the Audit Directive and WTA. Also note that, because of its specific objection to there not having been an audit firm level system of

quality control inspection, the judges expressed no opinion on the way the AFM had selected the statutory audit files to be inspected. They saw no need to do this. The selection method though is the central issue in the remaining sections of this paper.

5.2. The EY inspection by the AFM (AFM (2014))

In 2012 EY carried out 2640 statutory audits in the Netherlands (AFM (2014, p82). Note that AFM (2016) says 3300 statutory audits: 2640 will be used here. The AFM selected 10 files to inspect. It found 3 defective files. On the basis of this finding EY was fined (see AFM (2016)). It was fined for failing its duty of care towards its stakeholders.

This duty of care exists with regard to the audit work of all its engagement partners, not only the auditors involved in the 3 defective files. That is, the 3 defective files, in the eyes of the AFM, again without also assessing the EY system of quality control, signalled an EY wide system of quality control failure; manifested as an EY wide duty of care failure.

Two questions arise: (1) how did the AFM select the (sample of) 10 files inspected and (2) was this sample large enough. Note that AFM's inspection approach is reconstructed here. The AFM does not describe its approach in sufficient detail.

5.2.1 How: a reconstruction

To repeat: for 2012 financial statements EY performed 2640 statutory audits (the audit opinions were published in 2013). Hence, the 2640 statutory audits files are the population (182 PIE statutory audits and 2458 non-PIE statutory audits). There were 140 EY engagement partners in 2013 (AFM (2014, p.81)) to carry out these audits.

In a situation like this it will likely be costly to study the entire population in order to arrive at an opinion on whether EY has fulfilled its duty of care, i.e. has an adequate system of quality control in place. Of the 2640 statutory audit files handled by EY, the AFM, based on a list provided by EY of all its statutory audit clients, selected 10 files (and studied parts of these).

How did the AFM do this?

(a)

A straightforward approach would have been to draw a random sample from the 2640 audits. That is a sample of a large enough size in which all files (here: the 2640 statutory audits performed by EY) have an equal chance of ending up in the sample. A random sample (perhaps applying a form of 'monetary unit sampling') offers the possibility to draw quantitative conclusions about audit quality in the population of audits.

AFM (2014) does not use this approach. It does not explicitly explain why not.

(b)

The AFM explains the selection method it used instead as follows. It says (AFM (2016, p.131)):

'In selecting the statutory audits to be examined [we (the AFM) aim at]:

- variation in market segments (PIE audit clients and non-PIE audit clients)
- [variation in] audits by different external auditors of the audit firm;
- spread of audits over the organisational units of the audit firm, for example over branches and business units;
- the presence of an EQR;
- the presence of an IQR; and
- [variation in] average or higher risk profile (supplied by EY) of the audit.'

In fact, based on the explanation on p.131, AFM may have used even more criteria. It did not disclose which.

(c)

Continuing: the AFM may have selected in 2 ways. Again, it does not clearly explain which way. (1) It may have selected a portion of the EY 2640 audit files: the riskier portion (I thank Peter van der Heijden for pointing out this possibility). Or, (2) it may have selected precisely 10 (riskier) files directly. In both cases using the criteria listed.

(c1)

In the first case the AFM findings relate to the riskier portion of the work of EY's auditors. This interpretation will be true because it logically follows from the use of all its criteria, including the final criterion ('average or higher risk profile').

The reasoning of the AFM will likely have been that, if its opinion on the more risky files would be positive, it is plausible that the less risky files would also be fine. Note however that the AFM did not explicitly explain this line of reasoning in AFM (2014) and AFM (2016).

The AFM must have reasoned additionally that if the opinion on the riskier cases was negative the AFM did not need to consider investigating the lower risk cases. There is no evidence about the less risky files in AFM (2014) and AFM (2016).

Notice also, again, that in AFM (2016) a condemnation of the entire work of EY's auditors was made, not just the more risky files audits.

Assume that AFM indeed demarcated a risky portion of EY audit files, using its criteria. In that case, how did the AFM select the 10 files? Here again a random drawing of (the 10) files would have been possible. Here also a random sample (perhaps in the form of 'monetary unit sampling' using audit fees) would offer the possibility to draw quantitative conclusions about (quality in) the population.

(c2)

The alternative scenario is that the AFM arrived, using its criteria, directly on a sample of 10 audit files to be inspected from the 2640 performed in 2013, is problematic. That would constitute an extreme form of purposive sampling. The explanation on p.131 of AFM (2016) actually suggests that this is indeed how AFM selected the 10 files.

5.2.5. Reflections

Reflection 1

In AFM (2014) it is not clear that the 10 files were indeed drawn at random from the riskier files. To be able to quantify the consequence of not using a random sample assume that the AFM did draw the 10 files at random. In that case, there is random sampling from the riskier files.

Suppose, for example, that 200 files met the characteristics of high-risk files listed by the AFM. The number of 200 is not really important. There are 10 files randomly drawn. The goal is for these 10 files to be representative of the 200 files. As sample size increases it becomes more certain that the sample is representative.

So is this sample size of 10 large enough? It can be shown that 10 is in fact small. This makes an AFM conclusion '3 out of 10 defective: hence 30% of the EY higher risk files are defective' unreliable. It is too imprecise an estimate. It can be shown that a 95 % confidence interval runs from 4% to 56%. Because this interval is so large, the

conclusion of 30 % defective audit files for the population of riskier audit files, is not informative even if the 10 cases were drawn at random (which they were not).

In addition: (i) how large in fact is the population of riskier files (200 was assumed here)? The AFM states its criteria, but does not give the size of the riskier population. And (ii), because of the emphasis with which the AFM indicates (in AFM (2016)) that there is no statistical sampling, it is clear that the sample is not random. As a result, it cannot be ruled out in advance that the three files are exceptional for the work of EY's auditors, possibly even among the riskier files. The AFM is therefore unable to arrive at an opinion on the entire work of EY's auditors in respect of riskier files in 2012 on the basis of these 10 high-risk files.

Reflection 2

Returning to the scenario where AFM selected the 10 files to be inspected directly from the 2640 files using its criteria. This would constitute a form of purposive or judgement sampling. Again this would be a non-statistical sample which does not allow generalisation to the population of EY audits, for reasons already sketched. A further disadvantage of purposive or judgement sampling is that it creates (1) incentives to find an issue in the files selected by the selector (i.e. the AFM), and (2) it may suffer from hindsight bias. See Pecher, Solomon (2014) on the danger of hindsight bias in audit files inspections such as those of the PCAOB, and the AFM also, in general.

Reflection 3

The EY example problems readily generalises to all AFM PIE inspections: B4 and non B4 PIE audit firms, and to its (delegated) non-PIE audit firm inspections. Up until the AFM (2021) report the AFM has used this audit files inspection methodology. Therefore the negative appraisal here extends to all these inspections.

Reflection 4

The AFM approach to individual files inspection is used by all EU POBs. Although a number of these POBs have made part of the files selection random. But mostly the critique here extends to all these inspections as well. Therefore our negative appraisal extends to all these inspections.

Reflection 5

Additionally, given the AFM (2014) consequences, and those of the later inspections, note that the AFM explains in all these inspection reports that its negative files findings do not mean that the client financial statements concerned were defective. That is: the audit file is defective, but there is no audit failure. This raises doubts about the appropriateness of subsequent regulatory consequences.

Reflection 6

Also note, related to reflection 5, given the AFM (2014) results, and those of the later inspections reports, that the AFM could have brought the PIE audit firm engagement partners of defective audits to the Auditor Court. To date it has only done so once.

Reflection 7

By way of background, note that in the case of the 10 selected EY files in AFM (2014) EY performed 6 EQRs and 3 IQRs. For the 3 defective files there were 3 positive outcome EQRs (evidently) and 1 positive outcome IQR (there was no IQR for the other 2 files). What the AFM does therefore is not only evaluate the original engagement partner's with, but also that of the EQR partner and the IQR partner. The AFM does not comment on this.

Reflection 8

The AFM does not appear to inspect other engagements of an engagement auditor of a client file in which it finds defects. That underuses interesting available evidence.

Reflection 9

The AFM, in AFM (2014) and AFM (2016), appears to take the extreme position that a very small number of defective files (even only 1) signals insufficient duty of care on the part of the audit firms. That is, it uses in effect a 'zero defects allowed' approach. This appears to run counter to the 2006 Auditing Directive intention.

Reflection 10

In Vonnis (2017) the court accepted the position of the AFM that its selection of 10 files was not a 'sample'. Hence, a statistical critique (like that in this paper) cannot be used to argue with the AFM about the inspections outcomes. This is a problematic opinion.

Reflection 11

As explained earlier, the AFM in AFM (2021) changed its files selection approach. It now samples from the IQRs performed by the PIE audit firms inspected. It no longer uses the list of criteria it used earlier. Therefore it now deviates from the IFIAR approach. There is no explanation as to why this happened. Is this in any way linked to the EY and PwC court cases ?

Reflection 12

The AFM has also so far not disclosed in sufficient detail what capacity it has to do files inspections. The small number of inspections it performs may reflect resource limitations. Some insights are available. EY itself carried out 70 IQRs in the same year that the AFM inspected the 10 EY audit files (EY (2014)). It appears that EY itself has more resources available for audit quality inspections than the AFM, or is a far more efficient inspector. Aobdia (2018) mentions that one week of PCAOB time is on average required for an audit file inspection. For IQRs (in a US data based paper) Bell et al (2015, US data) mention an average IQR file inspection duration of 52 hours: slightly more than one week. Hanlon, Shroff (2022), in a survey study of audit firm POBs worldwide, found that in 2019 the AFM had 22 fte audit firms inspectors. This is in line with the between 10 and 25 AFM inspectors found and mentioned in Yuan, Florou (2021). Using these numbers and assuming that AFM inspectors work 1680 hours per year (a normal Dutch working year) the 22 AFM inspectors could inspect around 700 files in a given year. The number of audit files inspected annually since 2010 by the AFM is far lower.

6. Recommendations for 'POBs in the EU' on audit files inspections

(a) EU audit sector POBs audit files inspections started in the years after the 2006 Auditing Directive. IFIAR, coordinating statutory audit firms POB inspections, started operating in 2006 as well. The PCAOB started inspecting audit firms a few years earlier.

One observation is how unperturbed the audit sector POBs are by direct, theory based or empirical findings, based criticism, like those in the previous section, of their inspection approaches. There is never, in publications or on their websites, a reaction to these criticisms. An early example is the Peecher, Solomon (2014) powerful brief critical note (directed at the PCAOB). For critical remarks about PCAOB, also see: Houston, Stefaniak (2013) and Christensen (forthcoming). In the Netherlands, more recently CTA (2020) (see also Wallage et al (2020)) voiced strong criticism, along the lines of a number of the Reflections above, of the AFM audit files inspections approach. There is no AFM reaction on record. There is some movement on the part of some POBs (however without acknowledging the criticism). The Belgian audit sector POB, which only recently started its oversight in earnest, states that part of its audit files selection in Belgian statutory audit firms is done randomly. Also, the PCAOB recently said it 'will be selecting audits for inspection more randomly' (Maurer (2020)).

The recommendation here is that the POBs and IFIAR respond in a serious manner to criticism, or better still, actively organise criticism in the form of independent

advisory councils. An alternative could also be for the POBs to have a Chief Scientific Officer (Buijink (2016)).

In such an arrangement (advisory council en CSO being present) the POBs could/should also begin to triangulate their audit firm quality findings with other pertinent academic research findings. For instance, and directly relevant for the AFM: Bouwens et al (2011) and Peters (2020) find that financial reporting quality of Dutch listed firms increased (!) over the time period in which the AFM carried out its critical audit firm inspections. Buijink (2021), in a literature review, finds that Dutch financial reporting quality is high (!) relative to its neighbouring countries. AFM does not appear to use such information, nor do POBs elsewhere.

(b) Indeed, in these advisory councils academia will need to play a role as. For instance with regard to the following issue. Above, random sampling of audit files to be inspected is used to benchmark the AFMs approach. But on second thought the use of a random sample (correct size) from an inspected audit firm's audit files population might not be straightforward. The parties involved (POB, Audit firms and engagement partners) are economic agents. So pure random sampling may not be optimal or necessary. How to do files inspections properly in that setting ? There are a number of approaches suggested in the literature: see e.g. Lazear (2006) and Varas et al. (2020). For instance, with regard to the 'inspection rhythm, the AFM inspects the PIE statutory audit firms once every three years. That creates predictability. The NBA and the SRA inspect a sample of non-PIE firms every year in a six-year cycle. This creates an unpredictability of the inspections of the non-PIE audit firms, that does not exist for the PIE audit firms. AFM (and the other POBs)

could have long enlisted informed researchers to help it solve issues like this. There are no clear no sign they ever did, except perhaps for the PCAOB.

(c) POBs should also learn from IQRs inside the PIE audit firms. It was mentioned earlier that in the EY case, EY in the same period examined (2013) internally carried out 70 audit files inspections. Compared with the 10 AFM inspections. What does this say about the efficiency of the AFM inspections? Similarly, in 2019 the PCAOB sought comments on whether it should require selection of engagements for IQRs 'to include either random selection or an element of unpredictability' (PCAOB 2019). Evidently insights gained as a consequence of this call for comments will be relevant for POB inspections, and Advisory Councils or CSOs when in place, as well.

(d) Further issues the POBs (and Advisory Councils or CSOs) ought to study are the following. Consider radically different files inspection sampling approaches. The POBs could focus on the audit sector as a whole. I.e. sample from all statutory audits in a given year, or from all engagement partners in a given year (similar to audit firm IQRs). POBs could also screen statutory audit firm clients' financial statements and then sample from the audits of the more problematic (lower quality) financial statements. Also: consider the 'second' and 'third' sampling issues in addition to the primary sampling problem of sampling audit files. The AFM, as do other POBs, now 'sample' topics in audit files. They do not study the complete audit file. Moreover, for non-PIE statutory audit firms there is the problem of sampling in each year of the six year cycle from these audit firms. The AFMs, shows no sign, in publications or on their website, of seriously studying these other sampling issues. Nor do the other POBs.

(e) Following up on these recommendations (involving triangulation and academic advisory councils) would moreover help accounting and auditing academia in their teaching about audit sector POBs. The POBs and their oversight are currently a 'black box', so their *raison d'être*, operations and outcomes are very difficult to teach about (in the first place to aspiring auditing practitioners). This is problematic.

8. Conclusion

(1) Focusing on the audit files inspections of statutory audit firms POB in the EU, this paper demonstrates that the POBs approach is flawed: it has many disadvantages and no compensating advantages. Their approach to sampling (1) audit firms in the case of non-PIE audit firms, (2) audit files for all statutory audit firms, as well as (3) areas within the audits is problematic. The sampling approach used, very close to purposive sampling, makes the sampling results uninformative with regard to the quality of the audits carried out. This approach is advocated by IFIAR, hence the problems appear in all jurisdictions. A problem specific to the Netherlands is that Dutch Law in this area (WTA and BTA) differs in two small but significant ways from that which is stipulated in the Auditing Directive. This allows the AFM to uncouple inspections of audit firm quality control systems from audit files inspections. This uncoupling subsequently led to Dutch courts rescinding AFM fines. This paper bases its conclusions on a case study of the court cases related to the AFM fining EY (Netherlands). Note that although the judges rescinded the AFM fine, the courts failed to notice explicitly the discrepancy between WTA and the Auditing Directive just mentioned, nor did they object specifically to the sampling approach used by the AFM. This was a missed opportunity.

(2) A further problem with the POB statutory audit firm inspections is that they are not transparent. Outside stakeholders, including those in academia, are forced to reconstruct what the POBs do and they cannot be certain of the veracity of their reconstruction. Moreover it is not at all clear whether the files inspections are carried out efficiently. The POBs do not give information about their uses of resources available. In the EU all of this violates article 32: 6 of the Auditing Directive (2006 and 2014): 'The system of Public oversight shall be transparent.'. This problem also exists elsewhere, for example in the case of the PCAOB. Aobdia et al (2019), Eutsler (2020) and Christensen et al (2021) are recent papers that attempt to 'reengineer' and evaluate the PCAOB 'within audit firms audit files selection' algorithm. It is not clear on what grounds these algorithms must be confidential.

(3) This paper also argues that the AFM, and also the 'author' of legislation for the audit sector, the Ministry of Finance, should have enlisted informed researchers to help it solve the issues with AFM oversight highlighted in this paper from early on. They did not. Only recently did the Ministry of Finance create the CTA (CTA (2020) to inform the Ministry of Finance 'theory and empirical evidence based' about further steps in the regulation of the audit sector in the Netherlands. A further related, and broader, question one can ask is the following. Why does audit sector regulation exist at all, and why does it appear ever expanding? This is a legitimate and researchable question. In fact research exists in this area. A recent example is Bourveau (2021) where references are given to other relevant research. To this can be added the realisation of the fact that the number of audit failures everywhere is low compared to the large number of audits carried out. (on this, see Francis (2004), Ball

(2009) and empirically Hail et al (2017)). As explored in Bourveau (2021) this can in fact be expected on theoretical grounds. There are considerable incentives in place to make audit firms and engagement partners audit quality seeking, even in the absence of public oversight. The AFM and the Ministry of Finance appear not to be aware of this research.

(1), (2) and (3) lead to the conclusion that regulation of POBs of the audit sector, and the oversight itself, in the EU is not 'sound theory' and 'evidence' based.

(4) This conclusion leads to a final question: why does this state of affairs exist. Why is audit sector POB regulation in the EU not theory based, evidence informed and rational (see Pinker (2021) on rationality). To focus this question on the Netherlands: why is AFM regulation and oversight not rational: i.e. not Intellectually honest? It appears that insufficient incentives are in place that would cause AFM's statutory audit firm oversight to be rational. The Ministry of Finance and other stakeholders, e.g. the financial press, do not provide these incentives. A particularly striking example of the Ministry of Finance failing to do this is this proposal/explanation in the current Draft of further regulation of the audit sector in the Netherlands of the Dutch Ministry of Finance, in MvT (2021), p33): 'The present [proposal ... stipulates] that the quality control system of audit firms must ensure that serious deficiencies are prevented in multiple statutory audits. The AFM can then supervise this based on an examination [...] of a selection of audit files. With regard to the term "selection", it is considered that for the AFM to determine a violation of this provision, on the one hand, no representative or extensive sample or selection is required, but on the other hand, the selection must be of such a size that

it can indeed be legitimately argued that the deficiencies found reflect a deficient system of quality control. For example, about 10 statutory audits at a large public-sector audit firm; a smaller selection may suffice for an audit firm that conducts fewer statutory audits.'. This is the sampling method used by AFM (2014)! A second example is this remark from CTA (2020, lemma 16): 'In the opinion of the committee [CTA], the discussion about audit quality in the Netherlands has led to an urgent problem. Even if there is no quality problem by objective standards (if the quality of the audit was good), then the perception of a poor quality audit harms the public interest.'. This CTA attitude will evidently not discipline the AFM or the Ministry of Finance into becoming theory and evidence informed.

It appears that this conclusion also carries over to the regulatory and oversight activities of the IFIAR, the other audit sector POBs and the Audit sector Directive 'authors' in the European Commission.

(5) Note that POB's in the EU also inspect the quality of PIE financial statements directly. The AFM does this for Dutch Amsterdam Stock Exchange listed corporation. The AFM does not sample here: it studies all around 80 financial statements. In other European countries, e.g. the UK, the POB does sample here as well. Research findings related to sampled financial statement inspections in those countries are interesting also for the audit inspection findings discussed above. An example is: Florou et al (2020).

References

Accountancy Europe, (2018), Organisation of the Public Oversight of the audit profession in Europe: state of affairs after the implementation of the 2014 audit reform

AFM, (2010), Algemene bevindingen kwaliteit accountantscontrole en kwaliteitsbewaking

AFM, (2013a), Rapport naar aanleiding van AFM-onderzoek naar kwaliteit accountantscontrole en stelsel van kwaliteitsbeheersing en -bewaking bij negen PIE-vergunninghouders

AFM, (2013b), Themaonderzoek niet-OOB-accountantsorganisaties. Deel 1: NBA-kantoren; Deel 2: SRA kantoren; Deel 3: Samenvatting

AFM, (2014), Uitkomsten onderzoek kwaliteit wettelijke controles Big 4-accountantsorganisaties

AFM, (2015), Dashboard 2015: change and improvement measures: PIE audit firms

AFM, (2016), Boetebesluit EY

AFM, (2017), Kwaliteit PIE-accountantsorganisaties onderzocht: uitkomsten van onderzoeken naar de implementatie en borging van verandertrajecten bij de PIE-accountantsorganisaties en de kwaliteit van wettelijke controles bij de Big 4-accountantsorganisaties

AFM, (2019), Kwaliteit overige OOB-accountantsorganisaties onderzocht: uitkomsten van het onderzoek bij de overige PIE-accountantsorganisaties

AFM, (2021), Grip op kwaliteit: uitkomsten intern kwaliteitsonderzoek en kwaliteit wettelijke controles van OOB-accountantsorganisaties

Aobdia, D, (2018) The Impact of the PCAOB individual engagement inspection process: preliminary evidence, in: The Accounting Review, 93 (4), 53–80

Aobdia, D, et al, (2019), Which audits does the PCAOB monitor closely: evidence from the EDGAR database, SSRN

Ball, R., (2009), Market and political/regulatory perspectives on the recent accounting scandals, in: Journal of Accounting Research, 277-323

Bell, T., et al, (2015), Audit Firm Tenure, Non-Audit Services, and Internal Assessments of Audit Quality, in: Journal of Accounting Research, 461-509

Bourveau, T., et al, (2021), Public company auditing around the securities exchange act, SSRN

Bouwens, J., (2011), Rapport over de evaluatie van de wet toezicht financiële verslaggeving, Tilburg University

BTA, (2006), Besluit van 16 augustus 2006, houdende regels ter zake van de uitvoering van de Wet toezicht accountantsorganisaties

BTA, (2008), Besluit van 23 juni 2008 tot wijziging van het Besluit toezicht accountantsorganisaties

BTA, (2016), Besluit van 8 december 2016 tot wijziging van het Besluit toezicht accountantsorganisaties

Buijink, W., (2006), Evidence-based financial reporting regulation, in: Abacus, 296-301

Buijink, W., (2016), Nederlandse accountancy heeft geen probleem, in: Het Financieel Dagblad, 12-09-2016, 12-13

Buijink (2021), Why does almost everyone think that audit quality is problematic in The Netherlands: and why are they wrong ?, in: S. Van Puyvelde, C. Buts, Critical contributions to economics and beyond, Lexxion, 341-359

Carson, E., et al, (2021), Establishment of National Public Audit Oversight Boards and Audit Quality, 61p., SSRN

Christensen, B, et al, (2020, forthcoming), The PCAOB inspection process: a client-level analysis of a large firm's experience, in: *Auditing: A Journal of Practice & Theory*

Directive 2006/43/EC of the European Parliament and of the Council of 17 May 2006 on statutory audits of annual accounts and consolidated accounts

Directive 2014/56/EU of the European Parliament and of the Council of 16 April 2014 amending Directive 2006/43/EC on statutory audits of annual accounts and consolidated accounts

Eijkelenboom, E., M. van Opijnen, (2020), Artikel 14 Wta 'Zorgplicht': Een dode letter? Tijd voor een feniks herrijzenis!, in: *MAB*, 94, 187-198

EY Netherlands, (2014), Transparency Report

European Commission, (2008), Commission recommendation of 6 May 2008 on external quality assurance for statutory auditors and audit firms auditing public interest entities

Eutsler, J, (2020), Risky Business: Assessing the Generalizability of the PCAOB's Risk-Based Inspection Reports for Annually Inspected Audit Firms, in: *Auditing: A Journal of Practice & Theory*, 117–138

Florou, A., et al, (2020), Proactive financial reporting enforcement: audit fees and financial reporting quality effects, in: *The Accounting Review*, 167-197

Florou, A., Y. Shuai, (2021), The costs of public audit oversight: evidence from the EU, SSRN

Francis, J., (2004), What do we know about audit quality, in: *British Accounting Review*, 345-368

Hail, L., et al, (2018), Corporate scandals and regulation, in: *Journal of Accounting Research*, 617-671

Hanlon, M., N. Shroff, (2021), Insights into auditor public oversight boards: whether, how, and why they 'work', SSRN

Houston, R., C. Stefaniak, (2013), Audit Partner Perceptions of Post-Audit Review Mechanisms: An Examination of Internal Quality Reviews and PCAOB Inspections, in: Accounting Horizons, 23-49.

Leuz, C., (2018), Evidence-based policymaking: promise, challenges and opportunities for accounting and financial markets research, in: Accounting and Business Research, 582-608

MvT, (2021), Memorie van toelichting Wet Toekomst Accountancysector

Maurer, M., (2020), US watchdog will be selecting audits for inspection more randomly, in: Wallstreet Journal, 7-12-2020

NBA (2019), Factsheet

Offermans, M., A. Vanstraelen, Oversight and inspection of auditing, in: D. Hay, et al, (2014), The Routledge companion to auditing, Routledge, 179-187

Osma, B, et al, (2017), Public oversight systems for statutory auditors in the European Union, in: European Journal of Law and Economics, 517–552

Peecher, M., I. Solomon, (2014), PCAOB's 'audit failure' rate is highly suspect, in: CFO Magazine, February 27

Peters, C., (2021), Is de kwaliteit van de accountantscontrole gestegen?, in: MAB, 95, 169-178

Pinker, S., (2021), Rationality: what it is, why it seems scarce, why it matters, Allen Lane

Pouw, F, S. Maijor, (2008), Accountantstoezicht in toekomstig perspectief: van nationaal naar internationaal toezicht, in: MAB, 82, 108-117

Van Opijnen, M., (2012), Van collegiale toetsingen naar onafhankelijk toezicht, in: MAB, 86, 5-13

Varas, F, et al, (2020), Random inspections and periodic reviews: optimal dynamic monitoring, in: Review of Economic Studies, 2893-2937

Vonnis, (2017), Rechtbank Rotterdam, ECLI:NL:RBROT:2017:9977

Vonnis, (2019), College voor Beroep voor het bedrijfsleven, ECLI:NL:CBB:2019:235

Wallage P, et al, (2020), Het toezicht door de AFM: toezicht zoals toezicht bedoeld is? in: MAB, 94, 47-54

WTA (2006), Wet van 19 januari 2006, houdende het toezicht op accountantsorganisaties

WTA (2008), Wet van 12 juni 2008 houdende wijziging van de Wet toezicht accountantsorganisaties

WTA (2016), Wet van 12 oktober 2016 tot wijziging van de Wet toezicht accountantsorganisaties