

The value of screening tools in cartel cases

Malin Arve, Armando J. Garcia Pires, Ronny Gjendemsjø, Ignacio Herrera Anchustegui og Frode Skjeret

Prosjektet har mottatt midler fra det alminnelige prisreguleringsfondet.



The value of screening tools in cartel cases

Malin Arve, Armando J. Garcia Pires, Ronny Gjendemsjø, Ignacio Herrera Anchustegui, and Frode Skjeret

Abstract

In this paper, we analyze the value of screening tests for competition authorities in cartel cases to raise red flags about possible cartels and as a deterrence on new or existing cartels. We start by looking at one of the main problems of screen tests: false positives. We show that multiple screen tests can function as a useful tool to reduce false positives. We then study the legal foundations of screen tests considering the EU competition law and build a case to use screen tests to trigger dawn raids.

1. Introduction

One of the main duties of the competition authorities is to investigate the existence of cartels. The reason is that cartels are detrimental to consumers since they raise prices and therefore reduce consumer surplus. Cartels can also be harmful to competitors, especially if cartels increase barriers to entry, or reduce profitability of firms outside the cartel. In addition, cartels can reduce innovation in a market since they reduce competition.

It is important for competition authorities to have tools that help them identify cartels, the sooner the better, and if possible, that these tools help to deter the creation of new cartels or promote the dissolution of existing cartels. Traditionally, competition authorities have mostly relied on leniency (especially for whistle-blowers anonymous or not), usually by participants in the cartel or by competitors. For this reason, competition law in many jurisdictions has special provisions for whistle-blowers, like reducing the penalties for whistleblowing cartel participants.

Ideally, competition authorities would like to rely less on whistle-blowers and more on own resources to initiate cartel investigations. This would in principle allow competition authorities to identify cartels at an early stage and potentially have a deterrent effect on potential cartels.

One tool that can be used by competition authorities with this purpose is so-called screening tests. Screens are statistical tests done with the help of data from a given market that can raise the possibility of illegal behavior. Screen tests have been used for instance to identify tax fraud and evasion, credit card or insurance fraud, and insider trading.

However, screen tests are usually not considered as a proof of illegal behavior. Often the legislation does not allow for screen tests to be used as a proof or if they may be used, their value is very limited (as is the case in competition law, as we will see below).

One of the main obstacles for the legal use of screen tests, as with all statistical tests, is that the tests are prone to false positives and false negatives. Even so, screen tests have been used to create sufficient grounds to trigger a dawn raid or inspection over an undertaking(s) suspect of being engaged in illegal behavior. For example, in the case of insider-trading, financial regulators can trigger a dawn-raid after seeing trading of shares before important market information is made public. Still, even in more evident cases, it can be argued that this occurrence can be a false positive since the trading of shares was a mere coincidence. Unless a proof is presented in court (like communication between parts), a screen test is not enough for a conviction.

In this paper, we start by analyzing the legal value of screening tests for competition authorities in cartel cases. We also look to a major challenge associated with screens tests, false positives.¹ We then show ways for competition authorities to reduce the cases of false positives. We argue that the use of different screen test to identify collusion are more powerful to avoid false positives than the use of a single test. We proceed to discuss the legal foundations of screen tests in the light of the EU competition law, in particular, the requirement of showing overt communication to prove the existence of a cartel. In face of this, we build a case to use screen tests to trigger dawn raids. Dawn raids, in turn, can be used to obtain more data on screen tests and therefore to refine and fine-tune the screen tests for this market.

The rest of the paper is organized as follows. In section 2, we review the literature on screening tests for cartels. In section 3, we discuss the evidential value of screening tests in cartel cases. In section 4, we discuss the screening tools' value in the investigation of cartels, while we in section 5, look at screens as the ground for dawn raids. Section 6 concludes the paper.

2. Review of the screening tests literature

The literature on screening tools in competition policy has been growing steadily in the last decades. The reasons for this growth are twofold. First, screening tools have proved successful in many other areas when it comes to detecting illegal behavior, e.g. tax evasion,² credit-card

¹ False negatives are also problematic but, in our view, problems with false positives are larger since Competition Authorities waste resources (as is usually the case in cartel cases) looking into false positive cases, thereby reducing the resources available to go after real cartels. A false negative, leaves a cartel in operation in the present, but there is still the possibility that in the future it will be uncovered, without compromising the action of Competition Authorities to go after other cartels.

² M. Nigrini, 'A Taxpayer Compliance Application of Benford's Law' [1996] The Journal of the American Taxation Association 18, 72–91.

or insurance fraud,³ and insider trading.⁴ Second, competition authorities have actively been demanding tools that allow them to detect illegal behavior at early stages of cartel formation.

However, no gold standard for screening tools in cartel cases has materialized. This is partially due to the nature of screening tests: screening tests are statistical tools that raise the possibility of detecting illegal behavior but are not a proof of illegal behavior since all statistical tests are prone to error, i.e., false positives. In this sense, the main aim of screening tests is to raise red flags at an early stage to attract the attention of regulators (whether it is tax authorities, financial authorities, competition authorities) or other interested parties (like insurance firms, or credit card firms) so they can prevent, preclude, or gather evidence on illegal behavior.

Several factors have hindered the emergence of a gold standard of screening tests for collusion. First, most screening tools are not theory-based, and second, screening tools do not completely distinguish between competitive and collusive behavior, which contributes to false positives. For example, if some firms exit a given market causing prices to increase, is the price increase due to the exit of competitors (less competition) or a formation of a cartel by the remaining firms? As a result, while competition authorities have started to use screening tools to detect collusive behavior, there is no standard practice in what screening tool is the most adequate to every situation and market.

As we will extend on in the next sections, another obstacle for a wider use of screening tools is that screening tools are not covered by the competition legislation in most jurisdictions. In most countries, to prove the existence of collusion, competition authorities must present evidence of overt communication between economic actors. Obviously, this is not possible with screening tests since they are just statistical tools. As mentioned above, screening tests can only detect illegal behavior if statistical irregularities are found in the data.

Harrington⁵ concurs that the objective of collusion screens is not to provide legal evidence of collusion, but rather to detect potential illegal behavior at an early stage to pursue further inquiries that can lead to compelling evidence. In addition, Harrington⁶ also agrees that effective screening tools may have a deterrent impact on firms and hence, a preventive effect on cartels. For these reasons, economic analyses of collusive behavior using screening methods must aim to “screen industries as a matter of course; even where there is no hint of collusion”,

³ J. Harrington, ‘Detecting Cartels’ in Buccirossi, P. (ed.), *Handbook of Antitrust Economic*, (MIT Press. 2008).

⁴ C. Ewerhart; N. Cassola; S. Ejerskov and N. Vallad, ‘Manipulation in Money Markets’ [2007] *International Journal of Central Banking* 3, 113–148; C. Pirrong, ‘Detecting Manipulation in Futures Markets: The Ferruzzi Soybean Episode’ [2004] *American Law and Economics Review* 6, 28–71.

⁵ J. Harrington, ‘Detecting Cartels’ in W. P. Buccirossi’ (ed.), *Handbook of Antitrust Economics*, (The MIT Press 2008).

⁶ J. Harrington, ‘Detecting Cartels’ in W. P. Buccirossi’ (ed.), *Handbook of Antitrust Economics*, (The MIT Press 2008).

⁷ that is, ex ante screen analysis. Despite this, screening tests often rely on observations of prices or quantities in markets where cartels (or collusive behavior) have already been detected, i.e. ex-post analysis.

To make things more difficult, there is a wide range of screening tests in the literature. Some collusive markers focus on data from auctions, others on price data from retail markets. The screens for collusion in auctions and retail markets are different. Not only is the theory of bid-rigging cartels different from the theory of cartels in consumer markets, but also the type of firm behavior involved is different. In auction markets, firms compete for specific contracts at a specific point in time, while consumer markets are associated with competition for consumers on a regular basis. Furthermore, screens can be divided into complex and simple methods. Complex methods generally rely on econometric tools and structural estimations, while simpler methods tend to rely on measures of centrality and dispersion, like mean and coefficient of variation of prices. Complex methods require more resources and data.

The different screen measures available also rely in different assumptions regarding market shares, bidding data, prices and/or cost data.⁸ As a result, some authors propose screens based on mean and standard deviation of prices, while others on measures of kurtosis and skewness.⁹ Still other screens are based on mathematical laws of distribution of data such as Benford's law.¹⁰

Concerning screens in bidding (auction) markets, the screen literature says argues that auctions can be suspected of collusive behavior if, for instance, bids are highly correlated, or bids do not reflect costs in the market.¹¹ While screens using data on market shares, suspicion can arise if market shares are very stable or negatively correlated over time.¹²

⁷ J. Harrington, 'Detecting Cartels' in W. P. Buccirossi' (ed.), *Handbook of Antitrust Economics*, (The MIT Press 2008).

⁸ For a review see J. Harrington, 'Behavioral Screening and the Detection of Cartels' in C.D. Ehlermann and I. Atanasiu (eds), *European Competition Law Annual* (Hart Publishing 2006); L. Froeb, et al., 'Screening for Collusion as a Problem of Inference' in Roger D. Blair and D. Daniel Sokol (eds.), *The Oxford Handbook of International Antitrust Economics Volume 2*, (OUP 2014)

⁹ J. Harrington, 'Behavioral Screening and the Detection of Cartels' in C.D. Ehlermann and I. Atanasiu (eds), *European Competition Law Annual* (Hart Publishing 2006)

¹⁰ Benford's law describes the regularly occurring distribution of digits in data. If a given data set for a given market violates Benford's law, this can indicate collusive behavior; R. Abrantes-Metz.; S. Villas-Boas and G. Judge, 'Tracking the Labor Rate' [2011] *Applied Economics Letters* 18, 893–899; Giles, D., 'Benford's Law and Naturally Occurring Prices in Certain eBay Auctions' [2007] *Applied Economics Letters* 14, 157–161.

¹¹ P. Bajari, and L. Ye, 'Deciding Between Competition and Collusion' [2003] *The Review of Economics and Statistics* 85, 971–989; R. Porter and J. Zona, 'Detection of Bid Rigging in Procurement Auctions' [1993] *Journal of Political Economy* 101, 518–538; R. Porter and J. Zona, 'Ohio School Milk Markets: An Analysis of Bidding' [1999] *The RAND Journal of Economics* 30, 263–288.

¹² T. Bresnahan, 'Competition and Collusion in the American Automobile Industry: The 1955 Price War' [1987] *Journal of Industrial Economics* 35, 457–482; J. Hastings, 'Vertical Relationships and Competition in Retail Gasoline Markets: Empirical Evidence from Contract Changes in Southern California' [2004] *American Economic Review* 94, 317–328; J. Harrington, 'Behavioral Screening and the Detection of Cartels' in C.D. Ehlermann and I. Atanasiu (eds), *European Competition Law Annual* (Hart Publishing 2006)

The use of price and cost data in retail markets can demonstrate the existence of collusion if, for example, prices fail to reflect costs or reflect price-fixing behavior such as low variation in prices.¹³ For instance, higher prices accompanied by a reduction in variation of prices across consumers is one collusive price marker. Harrington argues that it has been documented that cartel formation in many instances is characterized by a series of price increases followed by large a price drop.¹⁴

Another important thing to note, is that price markers are partially constructed from real-world examples of cartel behavior (ex-post analysis) and rarely based on theoretical models of collusion. One reason for this is that the theory of implicit collusion generates very general hypotheses about firm behavior, which are difficult to translate into empirical measures of cartel detection. Another reason, as already remarked, is that it is difficult to distinguish between collusive and competitive behavior. As pointed out in Imhof *et al.*, however, two theoretical models can provide justification for the use of price screens.¹⁵ The first model is by Athey *et al.* who show that in an infinitely repeated Bertrand game where costs are private information (and vary over time), demand is inelastic, and if firms are sufficiently patient, optimal collusion is characterized by price rigidity.¹⁶ The second model is by Harrington and Chen, they show that prices are less responsive to cost shocks (prices are more rigid) in markets plagued by collusive behavior than in competitive markets in order to avoid cartel detection by regulators.¹⁷

Bajari and Ye develop a method to detect bid-rigging cartels ex ante based on first-bid sealed auction theory with asymmetric bidders.¹⁸ In the same line, Imhof develops an econometric

¹³ Y. Bolotova; J. Connor and D. Miller, 'The Impact of Collusion on Price Behavior: Empirical Results from Two Recent Cases' [2008] *International Journal of Industrial Organization* 26, 1290–1307; C. Crede, 'A Structural Break Cartel Screen for Dating and Detecting Collusion' [2018] *Review of Industrial Organization*, forthcoming; F. Esposito and M. Ferrero, 'Variance Screens for Detecting Collusion: An Application to Two Cartel Cases in Italy' [2006] Italian Competition Authority, Working Paper; J. Seaton and M. Waterson, 'Identifying and Characterising Price Leadership in British Supermarkets' [2013] *International Journal of Industrial Organization* 31, 392–403; For other examples of price data used to identify collusive behavior see M. Slade, 'Vancouver's Gasoline-Price Wars: An Empirical Exercise in Uncovering Supergame Strategies' [1992] *The Review of Economic Studies* 59, 257–276; S. Borenstein and A. Shepard, 'Dynamic Pricing in Retail Gasoline Markets' [1996] *The RAND Journal of Economics* 27, 429–451; J. Chevalier, A. Kashyap and P. Rossi, 'Why Don't Prices Rise During Periods of Peak Demand? Evidence from Scanner Data' [2003] *The American Economic Review* 93, 15–37; C. Lorenz, 'Screening Markets for Cartel Detection: Collusive Markers in the CFD Cartel-audit' [2008] *European Journal of Law and Economics* 26, 213–232; M. Lewis, 'Asymmetric Price Adjustment and Consumer Search: An Examination of the Retail Gasoline Market' [2011] *Journal of Economics & Management Strategy* 20, 409–449; M. Lewis and M. Noel, 'The Speed of Gasoline Price Response in Markets with and without Edgeworth Cycles' [2011] *Review of Economics and Statistics* 93, 672–682.

¹⁴ J. Harrington, 'Behavioral Screening and the Detection of Cartels' in C.D. Ehlermann and I. Atanasiu (eds), *European Competition Law Annual* (Hart Publishing 2006).

¹⁵ D. Imhof et al., 'Screening for Bid-Rigging: Does it Work?' [2017] CRESE Working Paper No. 2017–9.

¹⁶ S. Athey et al. 'Collusion and Price Rigidity' [2004] *Review of Economic Studies* 71, 317–349.

¹⁷ J. Harrington and J. Chen, 'Cartel Pricing Dynamics with Cost Variability and Endogenous Buyer Detection' [2006] *International Journal of Industrial Economics* 24, 1185–1212

¹⁸ P. Bajari and L. Ye, 'Deciding Between Competition and Collusion'. [2003] *The Review of Economics and Statistics* 85, 971–989

model to test for bid-rigging cartels.¹⁹ Imhof et al., screen for partial cartels in the context of bid-rigging cartels.²⁰ Partial cartels in bid rigging include collusion in auctions that does not involve all firms and/or all contracts. Recently, Huber and Imhof has also shown how machine learning can improve screening.²¹

The literature presented above illustrates both the lack of screening test that fits all environments and competitive situations, as well as the potential problems of such tools in the cases that they give rise to false positives.

3. The legal value of screening tests in cartel cases

The prohibition of anti-competitive agreements – cartels being the most prominent examples of them - in EU-law, Article 101 TFEU (**Treaty on the Functioning of the European Union**), consists of mainly two conditions. Firstly, there must be collaboration between firms in the form of an agreement, a concerted practice, or a decision by an association of undertakings. Secondly, the collaboration must have as its object or effect to restrict competition.²² The requirement of an agreement or a concerted practice is often said to be a requirement of overt communication. Therefore, one cannot establish a violation of Article 101 TFEU or other similar prohibitions, without evidence of overt communication.

The objective of a screening test is to predict whether there is collusive behavior or not. However, as noted above, screening tests are subject to uncertainty. This implies that the result of a screening test might indicate collusion even though no collusion occurs (false positive) or it might indicate no collusion although collusion occurs (false negative). In this article, we focus on false positives, since as argued in the introduction false positives may lead Competition Authorities to waste resources investigating cartels that never existed, using resources that could be spent on investigating real cartels.

Consider the hypothetical case when the screening test indicates the correct state of the world (collusion if there is collusion and no collusion if no collusion appears on the relevant market) with probability $p \in (0,1)$. If there is indeed collusion, the screening test will not indicate this with probability $1 - p$ (false negative). If there is no collusion, the screening test will nonetheless indicate that there is collusion (false positive) with this same probability. Depending on the statistical properties of the screening tests, these probabilities might of course differ, but for the sake of the argument in this article, we assume that they are the same.

¹⁹ D. Imhof, 'Econometric Tests to Detect Bid-rigging Cartels: Does it Work?' [2017] University of Fribourg Working paper SES N. 483.

²⁰ D. Imhof et al., 'Screening for Bid-Rigging: Does it Work?' [2017] CRESE Working Paper No. 2017-9.

²¹ M. Huber and D. Imhof, 'Machine learning with screens for detecting bid-rigging cartels' [2019] International Journal of Industrial Organization.

²² Most European national competition law consist of an identical or similar prohibition. In US antitrust law section 1 of the Sherman act contains similar conditions. Instead of agreements or concerted practices, the Sherman acts prohibits contracts or conspiracy, and the anti-competitive criterion is worded as restraint of trade or commerce.

The errors (or lack thereof) are related to the concepts of sensitivity and specificity and false positives. The sensitivity of a screening tool refers to the tools' capacity to detect collusion correctly in market-data when collusion takes place. The specificity of a screening tool refers to the tool's capacity to reject collusion in market-data when collusion never took place. For more details about the concepts of sensitivity and specificity, see appendix A²³.

Furthermore, screening tests are not able to distinguish between tacit and overt collusion. Tacit collusion, as pointed out by Harrington,²⁴ is easier to implement when prices in a market are very transparent, making direct communication between firms easier to avoid. Markets have become more transparent over time, partially due to digitalization and the internet era, thereby increasing the possibility of tacit collusion.

All the factors described above make screening tests insufficient as evidence of illegal collusion in themselves, and in this section, we will explain why it is so.

In EU competition law, the statement that establishing a violation requires evidence of overt communication is based on the interpretation of the conditions of agreements and concerted practices in the case law. The concept of an agreement is centered around the expression of a joint intention to behave in a certain manner on the market.²⁵ It is hard to imagine that someone expresses an intention without some form of overt communication.²⁶ When it comes to concerted practices the courts have interpreted this as consisting of direct or indirect contact between undertakings.²⁷ This also indicates that there must be some form of overt communication.

Any doubt relating to whether contact does not necessarily have to be overt communication is considered to have been removed by the judgment in *Wood Pulp*.²⁸ Here, the European Court of Justice (ECJ) stated that “parallel conduct cannot be regarded as furnishing proof of concertation unless concertation constitutes the only plausible explanation for such conduct”.²⁹ In this case, the ECJ reached the conclusion that an oligopolistic market structure

²³ Accordingly, to know the sensitivity and specificity of a screen test we would need to know the rates of false positives and false negatives of a given test. These rates are usually known for blood tests and other medicine tests, but that is not the case with the screens used in cartel cases. One reason for this is that there are many types of cartels, while instead all human beings are very similar making it easy to calculate these rates. Another reason is that we need many observations to calculate the rates of false positives and false negatives, but this is not available for most screen tests.

²⁴ J. Harrington, ‘Detecting Cartels’ in W. P. Buccirossi’ (ed.), *Handbook of Antitrust Economics*, (The MIT Press 2008).

²⁵ See e.g. case C-194/14 P *Treuhand* [2015], EU:C:2015:717, para 28, C-74/04 P *Volkswagen* [2006], EU:C:2006:460, paragraph 37 and joined cases C-2/01 and C-3/01 *Bayer* [2004], EU:C:2004:2, para 97.

²⁶ Note that the case law has accepted that vertical agreements may be based on tacit acceptance, as long as there is an invitation to behave in a certain manner, see *Bayer AG and Volkswagen*.

²⁷ See e.g. case C-49/92 P, *Anic*, ECLI:EU:C:1999:356 para 117 and case C-8/08, *T-Mobile*, ECLI:EU:C:2009:343, para 33

²⁸ Joined cases C-89/85 etc. *Wood Pulp* [1993], ECLI:EU:C:1993:120

²⁹ Joined cases C-89/85 etc. *Wood Pulp* [1993], ECLI:EU:C:1993:120, para 71

was considered an alternative explanation of parallel conduct. Consequently, parallel prices could not be considered as evidence of a concerted practice. This ruling by the ECJ is sensible and in line with economic theory. As already mentioned above, parallel prices are not necessarily evidence of collusion, and may be in fact be the result of (very transparent price) competition.

Even so, screening tests may indicate more than only parallel prices. For example, when (parallel) prices change, and the price changes cannot be explained by cost changes – or when prices remain constant despite cost reductions – this may be indicia of collusion. In such scenarios, the results of the screening test may indicate that the parallel prices are a result of collusion and not of competition.

From a legal standpoint, one could perhaps argue that in such scenarios, collusion is “the only plausible explanation of such conduct”, using the same language as the ECJ in *Wood Pulp*. This will rarely, if ever, be the case though, since screening tests, as we discuss above, are subject to some degree of statistical uncertainty and hence do not provide certain or sufficiently certain results to meet the legal requirements. Even if the test in a specific case would result in a solid and reliable result, the screening will not show if there is overt or tacit collusion. In this sense, the case law is quite clear when stating that to establish that there is an agreement or a concerted practice, some form of communication between the firms must be proven. This follows from the fact that the courts repeatedly having held that the requirement of contact between undertakings is supposed to ensure that Article 101 “does not deprive economic operators of the right to adapt themselves intelligently to the existing and anticipated conduct of their competitors”.³⁰ This has led to a consensus on Article 101 TFEU only prohibiting collusion based on *overt communication*, and not tacit collusion, which again lead to the requirement of proving overt collusion or contact between firms.

As demonstrated in this section, screens will not be sufficient as evidence to prove a cartel, since there is a need to prove overt communication between the firms. If there is evidence of overt communication about prices, price changes or market sharing, such communication will often be considered to be either agreements or concerted practice which have as its object to restrict competition, and there will not be a need to establish any effect on competition. On these grounds the result of screens of the market will not play a major role in establishing that there is a violation of the cartel prohibition, but as we return to below screening tools may play a significant role in the earlier stages of the investigation, either as a mechanism to detect cartels or a mean to establish sufficient grounds to perform dawn raids.

³⁰ See e.g. Case C-49/92 P *Anic* [1999], ECLI:EU:C:1999:356, para 116 and 117. Case T-69/04 *Schunk* [2008], ECLI:EU:T:2008:415 para 116. (emphasis added).

4. Screening tools and false positives

Considering that screening results may not be considered sufficient evidence of illegal collusion, the next question is how such screenings may be helpful to competition authorities. After all, the screening results may be an indication of collusion, and hence, the basis for a suspicion of illegal activity. For instance, can screen tests be used to trigger a dawn raid?

This question cannot be answered in a very straightforward way since screens are not perfect. It is therefore questionable whether screening tools are sufficient for “having reasonable grounds for suspecting infringement (...),”³¹ and therefore justifying a dawn raid. First, the range of conducts – tacit or not – causing collusive behavior is large. A screen – or a set of screens – must be able to detect illegal behavior along a range of dimensions: e.g. prices, quantities, market shares and more. Second, market characteristics may impose strict requirements on the quality of the screens. Thus, screens can only be expected to predict the true market conduct imperfectly.

As noted, screens are subject to uncertainty. In fact, the statistical properties related to various screening measures set stringent requirements on the sensitivity and specificity of the statistical tests used to avoid false positives (more on this in the appendix). Note that the problem of false positives is not restricted to screens, but to any statistical test that wants to capture an underlying phenomenon, such as for instance detecting a health condition in a blood test. The analysis in this section then applies not only to screens but to any binary statistical test that involves positives and negatives (see appendix).

The false positive paradox gives an example of screening for collusion where there are more false positive test results than true positive test results. Assume that a screen for collusion identifies collusive practices with very high precision. If this screen analyses many tenders per year, the number of tenders screened outweighs the high precision of the screen. Thus, the resulting list of screens indicating collusion when no collusion takes place will be higher than the number of screens indicating collusion when collusion takes place. This can illustrate two points relevant for competition policy.

First, the potential of screening tools to contribute to competition policy when it comes to sensitivity and specificity can vary from market to market. If a market is most likely characterized by non-collusive practices, the screening tools will have to be very strong to be effective and efficient tool for competition authorities. How much a market is prone to collusion is obviously not always known to the Competition Authorities. However, some rules of thumb can be taken from collusion theory, like for instance the number of actors in a market. Accordingly, markets with very few firms are more prone to collusion. National competition

³¹ Case T-325/16 *České dráhy v Commission* [2018], ECLI:EU:T:2018:368, para 36.

authorities could for example list markets according to the risk of collusion and run regularly screen tests on these markets.

Second, the requirements of screening tools regarding the sensitivity of screens, and its ability to be an efficient tool for competition authorities can in some markets be quite high. Accordingly, screening tools must have high degree of both sensitivity and specificity to be a “good” tool. Unfortunately, the literature on screening tools has not come that far yet to provide very robust screen tests.

Below we will illustrate the outcome of using screens to analyze a market when competition authorities use one to four screens to detect collusive behavior. We argue that combining several screening tools increases the usefulness of screening tools’ value in investigation of cartels.

Consider the following stylized example where the competition authority has x possible screening tools at its disposal and, for simplicity, they all have the same probability of indicating the true state and they are independent³².

As seen in Figure 1, adding more screens reduced the likelihood of false positives. Furthermore, the largest gains of running more screens occur from running just one screen or two screens. Running two screens substantially reduce the number of false positives. This is good news for Competition Authorities since running two screens is easier to routinely implement in screening possible cartels than three or more screens.

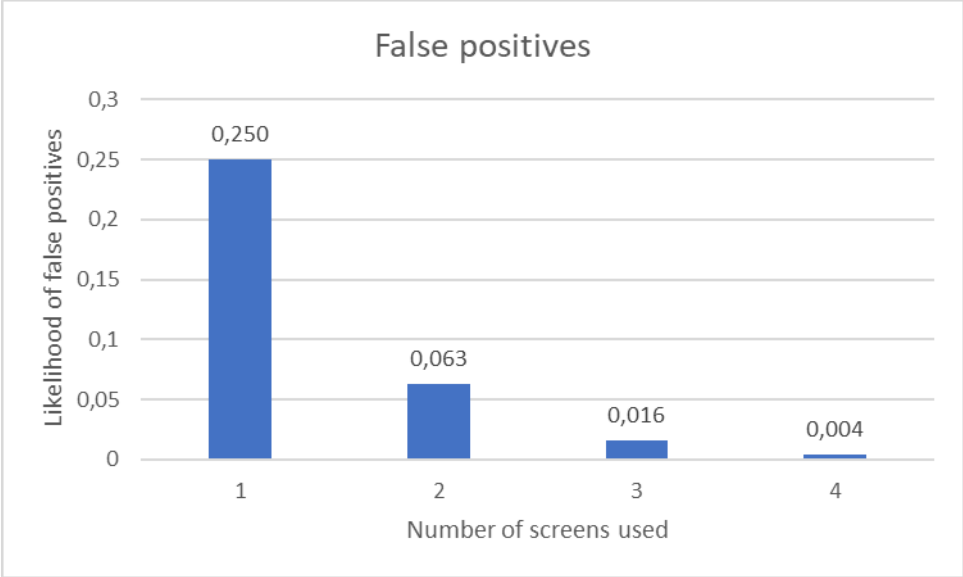


Figure 1. Likelihood of false positives for 1 to 4 screens.

³² In practice, screening tools are likely to be correlated as they are based on similar economic or statistical insight. As illustrated in further examples in Appendix B, correlation does not change the result.

5. Screens as the ground for dawn raids

Regulation 1/2003 sets in clear terms the powers granted to the Commission³³ and national authorities³⁴ while investigating possible infringements of competition law. The Regulation allows the authorities to request for evidence, such as demand information, as well as conduct inspections – or dawn raids. Both mechanisms would allow to collect information to develop the screening tool (as input data). They would thus increase the precision of the screening tests and reduce the likelihood and number of false positives.

First, the Commission and National Competition Authorities (NCA) have among their powers of investigation the right to request for information to undertakings or associations of undertakings as stipulated in Article 18 of Regulation 1/2003. Such information request may be used to gather information, directly from the involved undertakings, to confirm either the *indiciae* or suspicion arising from the screens or to construct screens with the available data – that in the case of collusion might not be of all involved parties. Requests for information need to provide what information is being requested as well as the time frame to do so and consequences of failing with such a request.

Additionally, and more importantly, Article 20 of Regulation 1/2003 grants the Commission the power to conduct “all necessary inspections of undertakings and associations of undertakings”. These powers exist even in cases in which an undertaking would not be under scrutiny by the NCA of its own Member State, as confirmed in the Orange judgment.³⁵ Unannounced inspections or dawn raids constitute an exceptional power of the Commission or NCAs in the course of a possible breach of competition law.³⁶ These powers are rather broad; so broad that some literature has put their legality into questions, particularly regarding the right to conduct inspections, sometimes, in an almost sarcastic manner, called ‘fishing expeditions’.³⁷ Also, due to their practical significance and value, unannounced inspections constitute the backbone of competition law investigations for uncovering possible infringements, in the case of abuse of dominance and related to anticompetitive agreements.³⁸ These inspections may involve entering into any premises, examine books, obtain copies of documents, seal premises, or conduct-recorded interviews, for example, and regardless of the medium in which this

³³ Regulation 1/2003 Recital (23 and 24).

³⁴ Regulation 1/2003 Recital (28).

³⁵ Case T-402/13 *Orange v Commission* [2014], ECLI:EU:T:2014:991.

³⁶ Riina Autio, ‘Explaining Dawn Raids: A Soft Law Perspective into European Competition Authorities’ Explanatory Notes on Unannounced Inspections’ [2020] 11 *Journal of European Competition Law & Practice* 9; Giacomo Di Federico, ‘Deutsche Bahn: What the Commission Can and Cannot do in Dawn Raids’ [2013] 5 *Journal of European Competition Law & Practice* 1, 29.

³⁷ Giacomo Di Federico, ‘Deutsche Bahn: What the Commission Can and Cannot do in Dawn Raids’ [2013] 5 *Journal of European Competition Law & Practice* 1, 29

³⁸ Lina Barauskaite, ‘Orange and EPIA e.a.: The Power of the European Commission in the Context of Inspections’ [2013] 6 *Journal of European Competition Law & Practice* 5, 327.

information is stored.³⁹ While a surprise for the involved undertakings, the Commission must notify well in advance the NCA in whose territory the inspection will take place.

Undertakings that are subject to a dawn raid are required “to submit to inspections ordered by decision of the Commission” (Article 20.4 of Regulation 1/2003), meaning that they are unable to refuse to such inspections. However, if the undertaking opposes the inspection, the Commission officers are to be assisted by national officers or even help in obtaining judicial authorization to conduct it. Failure to comply may lead to a fine up to 1% of the undertaking’s total turnover in the preceding year.

Inspections could be prompted by simulations and screening tests. However, inspections may be ordered only if a sufficient threshold of suspicion is reached. Article 20 of Regulation 1/2003 does not state a required standard or threshold of suspicion required to order an inspection. According to Article 20(1) the Commission may conduct all “necessary” inspections. The necessity is related to carrying out “the duties assigned to it”. Furthermore, Article 20(3) of the regulation requires the production of a “written authorization specifying the subject matter and purpose of the inspection”.

While the wording of Article 20 itself does not present a very clear standard to be met for conducting an inspection, there is some guidance in the case law. In *České dráhy* the General Court stated that an inspection decision “must be directed at gathering the necessary documentary evidence to check the actual existence and scope of a given factual and legal situation concerning which the Commission already possesses certain information, constituting reasonable grounds for suspecting an infringement of the competition rules.”⁴⁰ The General Court followed up this statement with saying that:

“...having reasonable grounds for suspecting infringement of the competition rules is a prerequisite for the Commission to order an inspection pursuant to Article 20(4)”.⁴¹

The prerequisite is meant as a safeguard to the need for protection against arbitrary or disproportionate intervention by the public authorities in the sphere of private activities of any person.⁴² In this sense, inspections have often been said to be in a clash with the rights of defense as well as human rights.⁴³ Their legality in EU (competition) law has been conditioned. Inspections are an exception to the right of privacy as enshrined in Article 8 of the European

³⁹ EUROPEAN COMMISSION, Explanatory note on Commission inspections pursuant to Article 20(4) of Council Regulation No 1/2003, para. 9.

⁴⁰ Case T-325/16 *České dráhy v Commission* [2018], ECLI:EU:T:2018:368, para 35

⁴¹ Case T-325/16 *České dráhy v Commission* [2018], ECLI:EU:T:2018:368, para 36.

⁴² Case T-325/16 *České dráhy v Commission* [2018], ECLI:EU:T:2018:368, para 34.

⁴³ G. Di Federico, ‘The Impact of the Lisbon Treaty on EU Antitrust Enforcement: Enhancing Procedural Guarantees Through Article 6 TEU’ [2010] *Il Diritto dell’Unione Europea* 805–833.

Convention of Human Rights and Articles 7 and 52 of the Charter of Fundamental Rights of the European Union.

However, they are justified within the scope of the right of effective judicial protection in Article 47 of the Charter of Fundamental Rights of the European Union and Article 6 of the European Convention of Human Rights as duly justified by the prevention of disorder and crime within the limits imposed by the rule of law in a democratic society.⁴⁴ As remarked by the General Court, “in order to justify inspections, it is not necessary for the information in the Commission’s possession to be of such a kind as to establish beyond reasonable doubt the existence of the infringement found in the contested decision”.⁴⁵ In other words, serious suspicions are sufficient, while a higher standard of proof is only required for finding a breach of competition law.⁴⁶ It is established case law that such obligation to justify and specify the subject-matter and purpose of the investigation is a necessary requirement for the Commission to conduct the raid.⁴⁷

The prerequisite of having “reasonable grounds” for conducting inspections is also found in Directive 2019/1, which harmonizes the powers of NCAs. This is not the case for EEA countries, like Norway, as the national enforcement of EEA competition law is not recognized by the EU, which results in the Directive lacking EEA relevance,⁴⁸ among these that of conducting inspections⁴⁹ investigations applying parallel national and EU competition law, or national law individually.⁵⁰ NCAs are empowered by Article 6 of the Directive in a similar manner as the Commission under Regulation 1/2003 to conduct unannounced inspections whenever “they can show that there are reasonable grounds for suspecting an infringement of Article 101 or 102 TFEU”,⁵¹ further clarifying the level of *indiciae* needed to conduct these.

⁴⁴ *Vinci Construction et GTM Génie Civil et Services v France*, CE:ECHR:2015:0402JUD006362910, para 63–65, 71–72 and 74. Now, with the negotiations for the accession of the European Union to the ECHR the relevance of the case law of the ECHR becomes all the more prevalent and, upon final accession, undertakings will be able to bring forth cases to the ECHR against the Commission’s inspection decisions. <https://www.coe.int/en/web/human-rights-intergovernmental-cooperation/accession-of-the-european-union-to-the-european-convention-on-human-rights>

⁴⁵ Case T-325/16 *České dráhy v Commission* [2018], ECLI:EU:T:2018:368 para 66.

⁴⁶ Case T-325/16 *České dráhy v Commission* [2018], ECLI:EU:T:2018:368, para 34. See also Emily Xueref-Poviac, ‘České dráhy, a.s. v European Commission: legality of Commission’s dawn raid’ [2018] 9 *Journal of European Competition Law & Practice* 9.

⁴⁷ Case 85/87 *Dow Benelux v Commission* [1989], ECLI:EU:C:1989:379, para 40. See also more recently: Case C-439/11 P *Ziegler SA v Commission* [2013], ECLI:EU:C:2013:513.

⁴⁸ The lack of recognition from EU of national enforcement of EEA competition law is further discussed in Ronny Gjendemsjø, ‘Article 56*’, in Arnesen et al (eds.), *Agreement on the European Economic Area: A Commentary*, 2018, Nomos/Hart, pp. 558-560; C.N.K.Franklin, H.H. Fredriksen and I.M.H. Barlund, *Norwegian national report for the 2016 FIDE Congress*, p. 371; R. Gjendemsjø ‘Norwegian Perspectives on EEA Competition Law Challenges’, *Europarecht*, Beiheft 1 2020, pp. 308-311.

⁴⁹ Directive (EU) 2019/1 of the European Parliament and of the Council of 11 December 2018 to empower the competition authorities of the Member States to be more effective enforcers and to ensure the proper functioning of the internal market

⁵⁰ Directive 2019/1 Recital 4.

⁵¹ Regulation 1/2019 Recital 31.

On this basis, the results of the screening tests, in themselves or together with other evidence, must be sufficient to give the competition authority reasonable grounds for suspecting a competition law infringement so that an inspection may be conducted. Considering that there are different methods to screen markets, and that the results of such screens may vary from case to case, it is not possible to present a clear answer on when they will lead to sufficient suspicion of illegal behavior to meet the prerequisite of “reasonable grounds”. However, some general remarks can be made. First, the threshold “reasonable grounds” is not a high one. The purpose of the threshold is as mentioned above protection against arbitrary or disproportionate intervention by the public authorities in the sphere of private activities of any person. As shown above, the application of more than one screening tool will reduce the risk of false positive significantly. If two or more screening tools indicate collusion in a market, this should be sufficient to meet the threshold of “reasonable grounds”. Even though the results of the screens don’t differentiate between overt and tacit collusion, the fact that they indicate collusion should be sufficient to meet the threshold for inspections. There is no condition at this stage of the investigation to prove that there is overt communication. Second, the purpose of the inspection will often be to investigate if there has been overt communication. To require proof of overt communication as a condition for an investigation, would significantly reduce the possibilities of detecting and sanctioning cartels.

6. Conclusion

In this paper, we have analyzed the value of screening tools for identifying and dismantling cartels by competition authorities. We documented that different screen tests have been developed in the literature in the last decades. For instance, screens based on statistical measures, like mean, variance, standard deviation, skewness, and kurtosis. These measures are usually developed ad-hoc and only loosely based on theories of collusion. Most of the literature also apply these measures ex-post to known existing cartels. However, screen measures are more valuable if they also began to be used ex-ante, to identify cartels at an early stage. The use of screen tests ex-ante can also act as a deterrence to potential new cartels.

We have shown why screening tests are not sufficient evidence for illegal cartel behavior. First, the competition law does not consider screens as valid proof. Second, screen tests can be prone to false positives. We argue, however, that screening tests may amount to sufficient grounds for adopting a decision to conduct inspections (dawn raids). Furthermore, we showed that while the question of whether this standard is met is not subject to review by national courts in the state where the dawn raid takes place, it is subject to review by the EU Courts.

We have also illustrated numerically the problem of false positives and argued that the different screens should be used together to reduce the number of false positives. Furthermore, screening tools should always be supplemented with economic analysis of the market in question to eliminate other causes for high prices, like costs and macro-economic conditions. Accordingly, if price of raw materials or inflation in the economy are rising, it is normal that prices in a given market are also increasing, without the need of a cartel to be present in this

market. These are factors which competition authorities must consider when deciding whether the screens provide “reasonable grounds” to suspect illegal cartel behavior.

Appendix A: Sensitivity and Specificity

In this appendix, we explain in more detail the concepts of sensitivity and specificity when analyzed in conjunction with false positives. First, we define first the following terms:

True: Collusion did take place

False: Collusion did not take place

Negative: The screening tool was wrong

Positive: The screening tool was correct

The following matrix illustrates the possible cases:

		True Market Behavior	
		Collusion	No Collusion
Predicted Market Behavior	Screen Reports Collusion	True Positive	False Positive
	Screen Reports no Collusion	False Negative	True Negative

Table A1: Confusion matrix: screens predicted versus actual market behavior

The sensitivity of a screening tool refers to the tools' capacity to detect collusion correctly in market-data when collusion takes place. The sensitivity of screening tools is defined as the ratio of true positive observations and total factual positive cases of collusion. This term relates to the economic expression effectiveness. A screening tool indicating that collusion took place, when collusion occurred, is an effective tool for competition authorities. Hence, a screening tool that scores high along the sensitivity-dimension will contribute to a competition policy highly effective when it comes to detecting cartels.

The specificity of a screening tool refers to the tool's capacity to reject collusion in market-data when collusion never took place. The specificity of screening tools is defined as the ratio of true negative observations and total cases when collusion never took place. This term relates to the economic expression efficiency. A screening tool indicating that collusion did not take place, when collusion factually never occurred, will reduce the number of additional analyses undertaken by competition authorities. Hence, a screening tool that scores high along the specificity-dimension will contribute to lower costs for the competition authorities.

Appendix B: False Positives with screen measures that are correlated

In this appendix, we illustrate that the statistical result regarding the likelihood of false positives for 1 to 4 screens also hold when screens are correlated. To do so, we simulate two models. For both models we assume that the specificity of all the screening tools is 0,75. We assume that the different screens' precision is positively correlated. Figures B.1 and B.2 summarize the correlation matrixes for the two simulations. Figures B.3 and B.4 show the false positive rates for the two simulations. The screens are applied to 10.000 tenders. To obtain confidence intervals for the figures, the market outcomes were simulated 50 times.

$$corr = \begin{bmatrix} 1 & 0.85 & 0.75 & 0.65 \\ 0.85 & 1 & 0.85 & 0.75 \\ 0.75 & 0.85 & 1 & 0.85 \\ 0.65 & 0.75 & 0.85 & 1 \end{bmatrix}$$

Figure B.1: Correlation matrix for simulation 1

$$corr = \begin{bmatrix} 1 & 0.85 & 0.85 & 0.85 \\ 0.85 & 1 & 0.85 & 0.85 \\ 0.85 & 0.85 & 1 & 0.85 \\ 0.85 & 0.85 & 0.85 & 1 \end{bmatrix}$$

Figure B.2: Correlation matrix for simulation 2

We can see from figures B.3 and B.4 that, as in the example in the main text, increasing the number of screens, reduces the number of false positives. Furthermore, also as in the main text, the largest gains of running more screens occur from running just one screen or two screens. Not only does running two screens substantially reduce the number of false positives, but the difference is also statistically significant at 95% level. This confirms that the results in the uncorrelated case in the main text, holds also for correlated models.

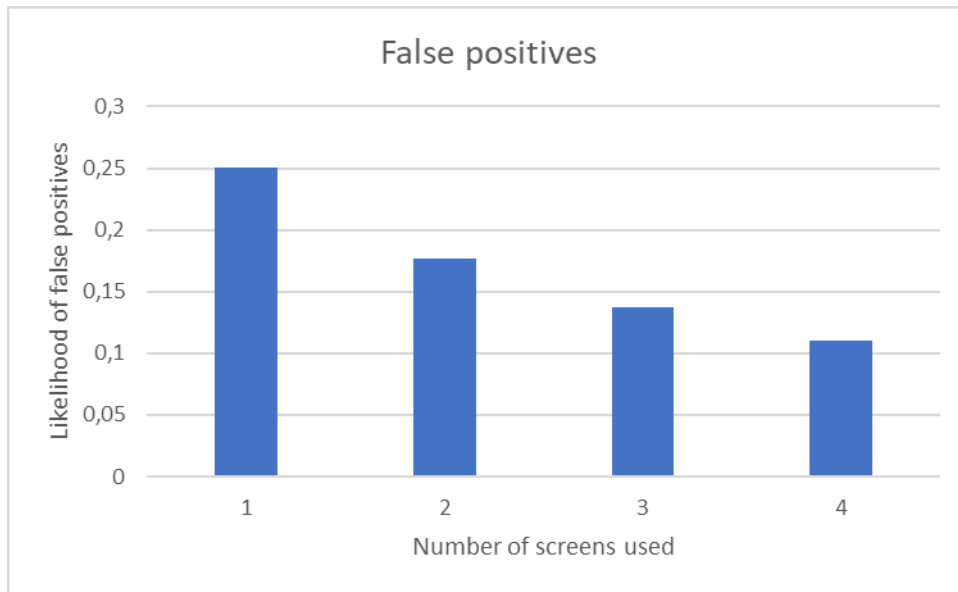


Figure B.3. Bootstrapped average false positives for 1, 2, 3 and 4 screens from simulation 1, i.e., using the correlation matrix in Figure B.1). 95% confidence interval based on 50 runs.

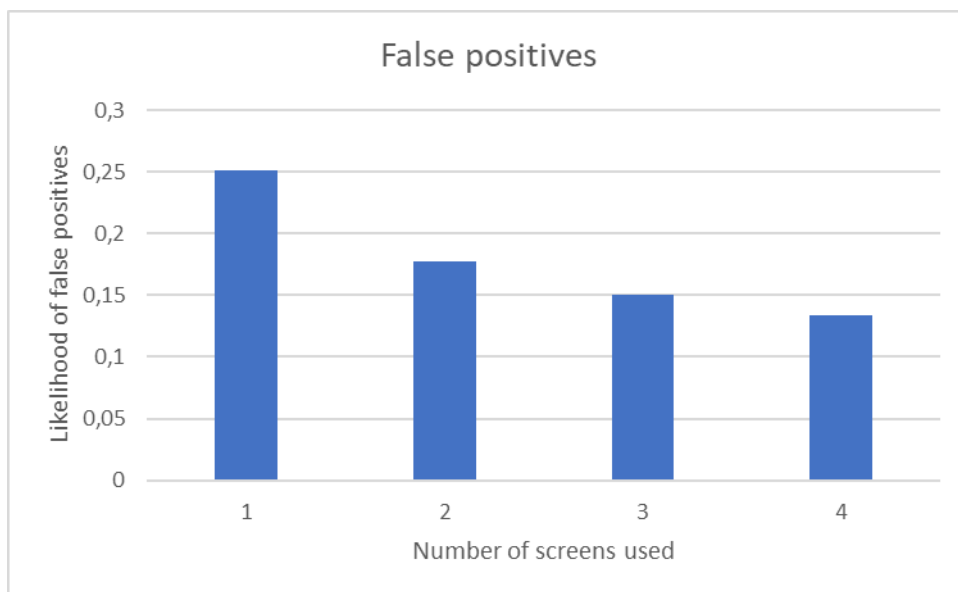


Figure B.4. Bootstrapped average false positives for 1, 2, 3 and 4 screens from simulation 1, i.e., using the correlation matrix in Figure B.2). 95% confidence interval based on 50 runs.