

The role of economics: testing for the extent of effects-based enforcement and its relation to the judicial review in different countries¹

April 2020

Y. Katsoulacos², S. Avdasheva³, K. Benetatou⁴, S. Golovanova⁵, G. Makri⁶

Abstract

This paper empirically investigates, for the first time to the best of our knowledge, the role of economics in Competition Authorities' (CAs) antitrust enforcement, not using the input measure of the number and qualification of economists contributing, but by constructing and measuring indicators capturing the extent and the type of economics used in the decisions reached and subjecting them to empirical analysis. We rely on information extracted from infringement antitrust decisions reached by the CAs of the EC, France, Greece and Russia.

Measuring the extent and type of economics used in decisions, allows us to identify the legal standards (LSs) adopted in assessing different conducts and their evolution over time and to undertake comparative analysis of the different LSs adopted in different jurisdictions. The LSs adopted can also be compared to their theoretically optimal level so we can capture the *quality of enforcement* by the extent of deviation from the optimum for each conduct type, the consistency in the application of LSs and the overall level of enforcement quality of the CAs. We show that on average, economic analysis has played a relatively modest role in investigations, though with significant differences between countries, with LSs closer to *Per Se* and no discernible evolution towards full effect-based, even when this would be the appropriate LS, except in the case of the EC for abuse of dominance cases. Overall, our empirical findings indicate *that there are significant difference in the overall level of quality of enforcement this been higher in EC than in France, both of these been significantly higher than in Greece and the latter significantly higher than in Russia.*

We also examine the relation between LSs adopted and the outcome of the judicial review of decisions appealed, i.e. the rate of decision annulment. We do not find consistent support for Neven's (2006) conjecture that the rate of annulment increases as LSs move closer to effects-based because "the scope for disagreement is greater when economic theory and evidence are important".

Key words: antitrust, economic analysis, legal standards, *Per Se*, effect-based enforcement.

¹ We are grateful, for financial support, to the National Research University Higher School of Economics Basic Research program and for comments from participants of the 13th CRESSE Conference, 29th June – 1st July 2018 and of seminars at the Toulouse School of Economics, Febr. 2018, CITY University, London, May 2019, ESSEC Business School, Oct. 2019, OECD Competition Division, Oct. 2019 and Dept. of Economics, Universidad de la Republica, Uruguay, Dec. 2019. We are particularly grateful for comments by Jacques Cremer, Pablo Ibanez Colomo, Antonio Copabanico, Xeni Dassiou, Antonio Gomez, Saqib Jaffarey, Frederic Jenny, Eduardo Ribeiro and Patrick Rey. Y Katsoulacos undertook research towards the latest version of this article in the Hellenic Observatory, European Institute of the London School of Economics where he was Visiting Professor in 2018 – 2019. He is grateful to the Institute for its hospitality. Of course, all errors and ambiguities remain solely our own responsibility.

² Professor of Economics, Athens University of Economics and Business. Email: yanniskatsoulacos@gmail.com

³ Professor of Economics, National Research University Higher School of Economics (HSE). Email: avdash@hse.ru

⁴ PhD, Vice-president, Hellenic Competition Commission.

⁵ Professor of Economics, National Research University HSE-Nizhny Novgorod. Email: svgolovanova@hse.ru

⁶ PhD candidate at Athens University of Economics and Business and case-handler, Hellenic Competition Commission.

1. Introduction and literature review

The last 25 or so years are widely considered as witnessing in many jurisdictions throughout the world a substantial increase in the role of economists and, though this view has not relied on much formal empirical backing, even in the extent and sophistication of economic analysis applied in the assessment of cases and in reaching decisions in competition law enforcement. A few countries such as the US and Canada are generally thought of as leading the way in this regard. But while this view, or, better, hypothesis, can be thought of as uncontroversial for merger control, it is far from uncontroversial for antitrust enforcement in many jurisdictions other than the US and Canada⁷. This paper contributes a comparative empirical investigation of the role of economics in the antitrust enforcement decisions of a number of countries

The extent to which the assessment of conducts in specific antitrust cases relies on economic analysis depends on the legal standard (or decision rule) that is used to make the assessment. The theoretical analysis of the choice of legal standards (abbreviated, henceforth, to LSs) by Competition Authorities (abbreviated, henceforth, to CAs) and Courts has been to a large extent normative analysis looking at the determination of optimal LSs, either from the point of view of error-cost minimisation (Easterbrook, 1984; Beckner and Salop, 1999) or welfare maximisation (Katsoulacos and Ulph, 2009). More recently there have also been contributions in the positive analysis of utility maximising CAs making LS choices that reflect their reputational and operational cost-minimisation concerns or, satisfying performance targets (Katsoulacos, 2019a; Avdasheva et.al, 2018, 2016).

These theoretical developments, have not however been accompanied by an equally significant effort to analyse, through detailed empirical research, what is the choice of LSs and the exact role of economics in antitrust decisions. International organisations like the OECD and others have been recording the rise of economists in CAs throughout the world⁸ trying to measure the role of economics from the input side. The much more important task of measuring the role of economics from the “output” side i.e. looking at the extent to which and how economics is used in the decisions reached by CAs has been until recently completely unexplored. It is easy to understand the reason: the latter requires a very large investment in the construction of data bases of the decisions reached by CAs over time and the examination of each one of these decisions to identify the exact role, extent and type of economic analysis utilised in reaching the decisions. However, having undertaken this investment, allows one, on the basis of an appropriate methodology such

⁷ And a few other jurisdictions in which an economics-based approach has been established for some years now, like the UK. For background discussions and empirical information concerning the use of economics in competition law enforcement see, among others, Baker (2003), Gavil (2008), Neven (2006), Schinkel (2008) and Lianos (2012).

⁸ The greater role for economists in CAs is reflected in the number of economists with advanced training employed by competition agencies. The EC has actively promoted the use of more economists in the DGCOMP. The tendency is clear: in the staff structure of the DGCOM, the share of economists has increased substantially during the past 10 to 15 years. For example, in 2017 the DGCOMP employed 190 economists; its Chief Competition Economist leads a team of 27 economists; the US Federal Trade Commission, employed about 70 Ph.D. economists (*The Handbook of Competition Economics 2018*, GCR 2018). The importance of economic analysis has been emphasized by international organization like the OECD and the UNCTAD. See e.g. Background Paper, OECD Policy Roundtable on Economic Evidence in Merger Analysis (2011); UNCTAD, The use of economic analysis in competition cases, Trade and Development Commission; OECD (2008), Presenting Complex Economic Theories to Judges. DAF/COMP(2008)3; also, the Speeches by Commissioner Kroes “Effective Competition Policy – a key Tool for Delivering the Lisbon Strategy”, Brussels, 3rd February 2005, and “Building a Competitive Europe – Competition Policy and the Relaunch of the Lisbon Strategy”, Milan, 7th February, 2005. Also, Gerber (2015), for the challenge facing developing countries / new jurisdictions and Shastitko (2018).

as the one described below⁹, to map the extent and type of economic analysis to the LS associated with this analysis and, hence, to identify the type and evolution of LSs adopted and the extent to which the LSs adopted diverge from the theoretically optimal LSs, and hence to determine the quality of enforcement.

In this paper we report the results of the empirical work that we have undertaken on the role of economics and the type of LS adopted, having constructed databases of infringement decisions made by four CAs: EC's DGCOMP, a mature European CA (the French CA), a smaller European CA with less resources (the Greek CA) and a recently established and modernised CA from BRICS (the Russian CA). For the EC, the database contains 170 antitrust infringement decisions for the years 1992 – 2016, for the French CA the database contains 295 antitrust infringement decisions for the years 2000 – 2017, for the Greek CA it contains 79 decisions for the years 1997 – 2017 and for the Russian CA it contains 987 decisions for the years 2008 - 2015¹⁰. The databases also contain information about whether these decisions were appealed and, most importantly, on the outcome of the appeal process following judicial review – i.e. information about whether or not the decisions were finally annulled by the Courts of Appeal.

Using the methodology of Katsoulacos, Avdasheva & Golovanova (2019), explained in summary below, the databases of infringement decisions have been used to map the extent and type of economic analysis, utilised in the assessment, to the LS associated with each decision. LSs are thought of as forming a *continuum* at the extremes of which are the *Per Se* (or object based) and the (“full”) *Effects-Based* (or rule of reason) standards. In this, we follow Jones and Kovacic (2017) according to which there is now a “recognition of an *analytical continuum* whose boundaries are set, respectively, by categorical rules of condemnation (per se illegality) or acquittal (per se legality) and an elaborate, fact-intensive assessment of reasonableness (Rule of Reason). *These poles are connected by a range of intermediate tests that seek to combine some of the clarity and economy of bright-line rules with the greater analytical accuracy that a fuller examination of evidence can produce*”. In some cases, whether or not additional tests will be undertaken depends on the outcome of the prior tests, such as, for example, under the “safe harbor” rule in the treatment of vertical restraints in the EU.

Our main objectives have been to:

- (i) Examine the LSs adopted in the assessment of different conduct categories (and the economic analysis utilised), by the four CAs investigated, their evolution over time, and the extent to which they diverge from their optimal level (see below).
- (ii) Propose and then measure, for each CA, indicators of the “quality of enforcement for each conduct category” (meaning, the quality of assessment of the decisions reached in each category), and of the “overall quality of enforcement of competition law” (CL) of each CA.
- (iii) Identify the consistency with which LSs for specific conduct categories are adopted and hence the extent to which the CA's choices contribute to legal certainty.
- (iv) Analyse the relationship between the judicial review process and the LSs adopted, specifically, analyse the relation between the latter and the outcome of the judicial review of decisions appealed, i.e. the rate of decision annulment.
- (v) Undertake a comparative analysis of the above for the four jurisdictions examined.

⁹ This methodology, proposed in Katsoulacos et.al. (2019), is based on the idea that there is a continuum of LSs between *Per Se* and rule-of-reason, an idea with a long history, reviewed briefly below.

¹⁰ The conduct types associated with these decisions will be described in detail below.

To measure quality, we take the optimal LSs to be the social welfare maximising LSs, taking into account a number of factors: decision errors, deterrence effects, administrability and legal uncertainty. What are these LSs? In a series of papers, Katsoulacos and Ulph (2009, 2011, 2015 and 2016) have attempted, by using a maximization-of-welfare framework to provide answers on how the factors above affect the choice of the (optimal) LS¹¹ and hence, indirectly, about the appropriate role and extent of economic analysis in CL enforcement. This analyses justify the use of a Per Se Illegality standard to treat hard-core horizontal agreements but also point quite strongly to the view that for a range of conducts, which now are understood not to be strongly presumptively illegal (that include vertical restraints and single firm exclusionary behavior)¹², and for which the developments in economic theory and modeling in the last 20 or so years improved significantly the discriminating quality of the assessment¹³, moving to assessment with effects-based standards will improve welfare due to a reduction in the costs of decision errors and an improvement in deterrence effects¹⁴. Our empirical results shed light on whether the LSs actually adopted are close to their socially optimal level or, there is a significant divergence and on how LSs differ between jurisdictions and over time. The view that there is a significant divergence, for the case of the EC was put forward, for example, by Geradin and Petit (2010, p. 31) who suggested, for EC abuse of dominance decisions, that the assessment has relied on «old, formalistic legal appraisal standards, and (has shown) a reluctance to endorse a modern economic approach»¹⁵. A similar view was put forward by Neven (2006) and, more recently in the book edited by Gifford and Kudrle (2015) and by Sokol D. (2017)¹⁶. This paper examines whether these views can be supported by detailed empirical evidence identifying the economic analysis used in decisions, how the EC compares to other CAs and whether there has been an evolution towards more effects-based LSs in more recent years in these CAs.

It is important to note here that the choice of LS is dependent also on the *substantive standard* adopted. While in academic discussions this is usually assumed to be welfarist (liability requiring a showing of adverse effects on welfare¹⁷), in practice this is often not the case. For example, the substantive standard may be just to “protect the economic freedom of market participants”, or, the pursuit of a “system of undistorted competition” (Wils, 2014), without obligation to show adverse effects on consumer welfare or efficiency (Rey and Venit, 2015) – which would imply that any conduct that puts one or more competitors at a disadvantage would

¹¹ Extensive references and reviews of the literature related to these issues are contained in these papers. See also J Padilla (2011), page 435.

¹² But which up to the 1990s were widely considered as strongly presumptively illegal. See for more details Katsoulacos, Avdasheva and Golovanova (2016).

¹³ That is, the ability of the assessment to discriminate accurately between harmful and benign conducts.

¹⁴ Which are likely to more than compensate for higher administrative costs and legal uncertainty. As Jones and Kovacic (2017, p 260) note «many jurisdictions apply a rule of per se illegality, or virtual per se illegality, against some horizontal agreements such a price fixing. The extent to which such a rule should be expanded beyond this.....is much more controversial and contested». As they indicate (p. 269) nowadays in the US, vertical restraints, mergers and single-firm exclusionary behavior are not assessed by per se. Also Blair and Sokol (2012).

¹⁵ Relying on the predictions of sound economic analysis has been stressed by the OECD not just in the context of developed countries but equally and perhaps more importantly in developing ones. For example, in its recent report evaluating the Russian competition authority, that has in the last few years become the largest CA in the world, the OECD (2013) makes as its top recommendation that it must “improve the quality of economic analysis and its application to competition enforcement throughout the CA and in support of improved judicial decisions”.

¹⁶ For a more detailed examination of just the EC case see also Katsoulacos and Makri (2020).

¹⁷ Consumer or total welfare – see also below.

be considered unlawful¹⁸, irrespective of the ultimate consequences of the conduct for welfare¹⁹. The link between substantive standards and the choice of LSs has been discussed recently and it has been demonstrated (Katsoulacos, 2019b) that adopting non-welfarist substantive standards increases the likelihood that, along the *continuum*, Per Se LSs are applied²⁰ and a limited amount of economic analysis is utilized.

In contrast to the theoretical analyses, systematic empirical assessments of LSs have been very limited. There are case-studies based on particular decisions, but there is no *statistical representation* of the LSs applied by competition authorities, as proposed in this paper. The absence of empirical measurement and statistics on LSs limits our ability to make international comparisons of LSs applied in different jurisdictions and makes general judgments on the role of economic analysis speculative and inexact. Further, it impedes the analysis of the evolution of LSs over time and the consistency with which they are used in assessing specific conduct types. And, it restricts our ability to investigate rigorously whether and how the extent of economic analysis used impacts on the judicial review process (the rate of annulment of CA decisions). As noted above, it is our objective in this paper to deal with all these questions.

We proceed as follows. First, our antitrust decision databases are used to construct LS indicators (LSIs) associated with each antitrust decision. We distinguish between six LSIs (between Strict Per Se and Full Effects Based²¹) depending on the extent and type of economic analysis utilised, as explained in the next section. Decisions are categorised into four main conduct types or groups: hard-core horizontal agreements, other horizontal agreements (such as information agreements) and concerted practices, vertical agreements and restraints and abuse of dominance practices. Next, we develop a number of indicators. We measure the Weighted Average Legal Standard (WALS) adopted for each conduct group (the weights been the share of each LS used in assessing decisions in each conduct group), and the degree of concentration (CONC) for the WALS of each conduct group. The higher is CONC the greater the concentration on specific LSs when assessing specific conducts and hence the greater the certainty with which it can be anticipated that a specific LS will be used in the future for a conduct. The standard deviation of LSI is another measure of the uncertainty that surrounds the anticipated LS for a conduct. We also develop *indices of quality (Q) of enforcement* measuring the extent of deviation of WALS from its optimal level for each conduct group relative to the maximum theoretical deviation, as well as an overall index of quality of enforcement by the CA across all conduct groups. We then provide evidence about how the probability of annulment is related to the LS adopted.

We show that, on average, economic analysis plays a relatively modest role in antitrust investigations, though with quite significant differences between countries, with LSs that are closer to *Per Se* and no discernible evolution over time towards full effect-based, for conduct groups for which this would be the appropriate LS, with the very important exception of the EC for the abuse of dominance conduct group. Also, inconsistency in the choice of LSs for specific conducts can, in

¹⁸ The meaning of “preserving undistorted competition” was actually made clear by the EU General Court which, upholding in its entirety the Commission’s 2014 Decision on *Intel*, argued that making it more difficult for a rival to compete “in itself suffices for a finding of infringement”.

¹⁹ Rey and Venit (2015) note that the effects-based standard *starts* with a showing of a distortion of the competitive process but, in order to assess this distortion and find liability, one “should (also) look at the actual or likely *effects of the conduct*“, on consumer welfare or efficiency (p. 17, italics ours). Some economists have been arguing for a total welfare standard, e.g. D. Carlton (2007). For a recent review of the recent debate see Katsoulacos, Metsiou and Ulph (2016). Also, CAs often take into account the presence of “public interest concerns” as additional liability criteria.

²⁰ Though, non-welfarist substantive standards are consistent with the use of effects-based LSs (Katsoulacos, 2019b).

²¹ This is broadly considered to be, to use the US terminology, the rule-of-reason LS, but see also Vickers (2005).

some cases, create legal uncertainty for firms, this been most serious in the case of Russia. Overall, our empirical findings indicate *medium to high quality of enforcement in terms of the appropriateness of the LSs adopted in the EC, quality been higher in the EC than in France, in France significantly higher than in Greece and in the latter significantly higher than in Russia.*

Our empirical investigation of the relation between LSs adopted and the outcome of the judicial review of decisions appealed, i.e. the rate of decision annulment, shows that we cannot find consistent support for Neven's (2006) conjecture that the adoption of effects-based LSs are associated with higher annulment rates because "the scope for disagreement is greater when economic theory and evidence are important", though, at least in some cases, when LSs increase beyond a certain point towards effects-based the annulment rate increases. However, this relation is not monotonic and, most importantly, it does not hold (indeed the opposite seems to be the case) in the EC, for the case of which the conjecture was originally expressed by Neven (2006). Our results, at least for Greece and Russia, suggest that the relation may be a U-shaped one²².

The remainder of the paper is organized as follows. The second section describes the methodological approach to measuring empirically LSs. The third section reviews the characteristics of the datasets of each one of the countries examined. The fourth section contains the empirical analysis undertaken and a detailed discussion of the results obtained related to the objectives described above. The fifth section concludes.

2. A methodology for identifying the extent of economic analysis and legal standards in antitrust enforcement

The methodology, first proposed in the paper of Katsoulacos et.al. (2018), begins with the assumption that there are variations in the LSs adopted in competition law enforcement, encapsulating the idea that it is best to think of LSs as forming a *continuum* at the extremes of which are the *Strict Per Se* (or object based) and the (full) *Effects-Based* (or rule-of-reason) standards. The progression towards LSs closer to full effects-based requires that additional blocks or components of economic analysis are applied. These are associated with the definition of the relevant market, the assessment of market power, the assessment of whether market power raising or exclusionary effects are present, the articulation of a theory of consumer harm, the assessment of efficiency effects and the assessment of what is, ultimately, the welfare impact of the conduct. These components can be identified by analyzing the documents on particular decisions made by a CA. Specifically, all the information could be extracted from the texts of the decisions and can be categorized by assigning to the variable corresponding to a component of economic analysis a value "Yes=1 (in case the analysis has been undertaken)" or "No=0 (otherwise)".

²² This is consistent with the predictions of a model of utility maximizing CAs, such as that of Katsoulacos (2019a). Then, the CAs may optimally adopt LSs in the assessment of some conducts that are lower (closer to Per Se) than those adopted by the Courts. This can happen even when the CA can predict with certainty the Courts' LSs, but it is more likely to hold when there is uncertainty about the latter – as is likely to have been the case in Russia and, to a lesser extent, Greece in the early part of the period examined (but is not likely to be the case in the mature and experienced jurisdiction of France and the EC). The Courts regard these LSs as "wrong" (or, "too low") and this makes the observed probability of annulment correspondingly high exactly when the observed LSs adopted by the CAs are low. The probability of annulment then falls when the CAs, in assessing other conducts, adopt the same ("correct") intermediate LSs as the Courts. In this model, when the CAs follow the Courts in adopting higher than intermediate (closer to effects-based LSs) the probability of annulment rises according to the Neven (2006) conjecture.

The methodology identifies four broad components or blocks of economic analysis (A, B, C and D in Table M1) that must be performed for the investigation of the CA to constitute a full effects-based (or rule-of-reason) analysis. One or more variables (sub-components) make up every one of these main components of economic analysis (e.g. sub-components C.1, C.2, C.3 make up analysis component C, etc). If the CA does not adopt a full effects-based LS then some of the sub-components will not be part of the assessment. We focus, for the purposes of this paper, on all antitrust conducts, other than the exploitative conducts²³.

Thus, the analysis components under consideration [to which we assign (0/1) scores] are:

Table M1: Types of economic analysis taken into account in the construction of the indicators of the extent of economic analysis

Block or component of analysis	Description	Comment	Score
A	Discussion of the nature and characteristics of the conduct	Since in all cases there must be some discussion of the nature and characteristics of the conduct, we should not get a score of “0” here – in this sense this category is not needed. It is included for purely formal reasons, to remind ourselves that an overall score of “1” is a strict <i>Per Se</i> approach to the assessment, meaning that the CA only considered the nature and characteristics of the conduct.	0/1
B	Market Analysis	Comment	
B.1	Basic analysis of market characteristics based on available information and market statistics	This is economic analysis that is necessary in order for a CA to contextualise a conduct. This typically involves information about industry structure, the firms, the structure of demand and the technology, determination of market shares (without formal analysis of market definition).	0/1
or B.2.	Formal market delineation and market share determination	Market definition based on more sophisticated economic tests (e.g. SSNIP test, Price correlation and Critical loss analysis).	
C	Evidence on restrictions of competition/harm imposed	Comment	
C.1.	Analysis undertaken to identify if conduct has market power	This need not include the construction of a formal model (e.g. examination of incentive compatibility constraints in a concerted practice case, or examination of how exclusive contracts could lead	0/1

²³ Our methodology assumes that the economic analysis components that must be included in the effects-based assessment of the specific (non-exploitative) conducts that belong in our broader conduct categories (defined below), are the same.

Block or component of analysis	Description	Comment	Score
	enhancing (e.g. through agreements) or exclusionary (e.g. in monopolization practices) effects	to exclusion or prevent entry in the specific context, or “equally efficient competitor test”). But it must indicate a serious effort to demonstrate the presence of such effects.	
C.2.	Articulation of theory of harm to consumer welfare (without taking into account of efficiencies)	When “scoring” CAs decisions this need not be a full-blown formal analysis; one could also score an effort towards determining whether exclusionary effects can be expected to create a negative impact on consumers (through a price increase or a reduction in quality or diversity).	0/1
C.3.	Analysis of potential efficiency effects expected to benefit consumers	Analysis should be based on efficiencies that are expected to result from the conduct, <i>that will create benefits to consumers</i> (again, this need not be very sophisticated but must indicate a serious effort to take efficiencies into account). Analysis of potential efficiency effects relating to factors that tend to prevent a price rise or other harm to consumers. NOTE: Counterfactual analysis ²⁴ may be undertaken under any of the C components – though this is not strictly necessary for considering the effect as established.	0/1
D	More effects-based analysis	Comment	
D	Balancing of potential anticompetitive effects of conduct with <i>all</i> the potential efficiencies and determination of the final impact on <i>total</i> welfare.	This is any analysis “over and above” the analysis that may have been included under “efficiencies” above (taking into account efficiencies that need not impact consumers, especially in the short-term). By “balancing” here we mean any formal economic analysis that attempts to measure the net effect of the conduct, that may or may not be related to efficiencies - e.g. balancing the short-term and long-term implications of refusal to license (or of compulsory licensing) an innovative activity.	0/1
Maximum Total Score		6 (when the Substantive Standard is total welfare)²⁵ 5 (when the Substantive Standard is consumer welfare)²⁶	

Source: Revised from Katsoulacos et.al. (2019)

²⁴ I.e. analysis proposing that the theory of harm is not valid and demonstrating the absence of foreclosure effects and consumer harm of an exclusionary conduct.

²⁵ That is, when the criterion for liability is whether or not there is an adverse effect on total welfare.

²⁶ That is, when the criterion for liability is consumer welfare. Analysis component D need not be undertaken if the substantive standard is that of consumer welfare.

Note that the value (1 or 0) of an analysis variable (e.g. of B.2 or C.2 etc) is based on a judgment whether the relevant analysis has been undertaken or not and it says nothing about the correctness or quality of the analysis or of the data used. In other words, the value of an analysis variable indicates *whether the competition authority, in the particular case, has tried to address the specific question associated with that analysis variable.*

Effects-based scores and types of legal standards

On the basis of the above methodology, one can construct *effects based scores* (EBS), using the 6 analysis components above. EBS are calculated as the sum of the analysis variables presented in Table M1 – with a minimum of 1²⁷ and a maximum of 6. The question is: *Is it reasonable for undertaking empirical analysis to use data extracted from decisions that aggregate scores over many different conducts (e.g. all the non-exploitative antitrust conducts)?*

The answer is that a straight aggregation of scores across different conduct types will not provide indicators which can be used to undertake meaningful empirical analysis of the extent of economic analysis and type of LSs adopted. Such aggregate indicators cannot be used to measure meaningfully whether economic analysis is used optimally – since optimal LSs can only be defined at the level of each conduct. Furthermore, such aggregate indicators cannot be used to make comparisons between different countries and over time – since the level of the aggregate indicator will depend on the composition of conduct types that will be different for different countries and will change over time, so, for example, an EBS indicator score of, say, 2,91 for both Greece and France certainly does NOT mean that the extent of economic analysis relative to some optimal level is the same in Greece and France given that the composition of conduct types may well be completely different between the two countries²⁸. Moreover, such aggregate indicators cannot be used to examine how changes in the economic analysis, if measured by changes in the value of the aggregate indicator, affect the annulment rate of the CA's decisions, since the latter is expected to be influenced by what type of economic analysis²⁹ is utilized and how this changes, while a given value of the aggregate indicator cannot reflect what types of analysis are utilized and, when the value of the indicator changes, what type of economic analysis is responsible for the change in the indicator's value.

To avoid these problems we construct Legal Standard Indicators (LSI) from the information on the scores received on each of the analysis components in Table 1 for each decision, as follows. For the set of decisions (say, S1) in which only analysis component A is undertaken we say that the LSI is equal to 1. For the set of decisions (S2) in which analysis components A and B are undertaken we say that the LSI is equal to 2. For the set of decisions (S3) in which analysis components A and B and C.1 are undertaken we say that the LSI is equal to 3, and so on. Now, by comparing the different sets of decisions we can identify the effects of *additional* economic analysis. For example, by comparing decisions in S2 with decisions in S3 we can identify the effect of adding the analysis component C.1; by comparing decisions in S3 with decisions in S4 we can identify the effect of adding the analysis component C.2. We are also able to identify the frequency

²⁷ There must always be some discussion at least of the nature and the characteristics of the conduct so the minimum value of EBS is 1.

²⁸ E.g. in France there may be proportionally many more decisions on conduct types for which the appropriate LS is Per Se or close to Per Se.

²⁹ For example, different types of economic analysis can lead to a score 3 and different ways of increasing economic analysis can increase the score from 3 to 4 but the implications of each case for the rate of annulment may not be the same.

with which the CA applies the analysis associated with each one of the sets in assessing different conduct types and, hence, infer the extent to which the CA favors a certain legal standard for the different conduct types (see below).

Katsoulacos (2019a) and Katsoulacos et al. (2019) distinguish between a number of distinct LSs between Strict Per Se and Full Effects-Based, corresponding to the above mentioned (sets) and LSIs (see Table M2 below). A brief description of the main LSs follows.

Under the *Strict Per Se (SPS)* LS the CA makes decisions on the basis only of the purely formal characteristics of the conduct under investigation, relying on strong presumptions about the implications of the general class of conducts to which the specific conduct belongs for welfare. Alternatively, one can say that under SPS LS the CA makes inferences about effects (on welfare) from the formal characteristics of the conduct.

The *Modified Per Se (MPS)* LS can be considered as a Per Se rule subject to a Significant Market Power requirement or, more generally, as supplementing Per Se by undertaking analysis of market characteristics, for example, when assessing conducts under abuse of dominance or in an information exchange agreement or in a concerted practice for which there is no strong hard evidence of collusion. Alternatively, one can say that under MPS LS the CA makes inferences about effects (on welfare) from the formal characteristics of the conduct, detailed analysis of market characteristics and, depending on the type of conduct, the implications of these on incentives for achieving sustainable collusion and/or on the size of extant market power³⁰.

Truncated Effects Based (TEB) is a higher LS, under which decisions about whether or not there is liability in the case of a specific conduct are reached by establishing that the characteristics of the specific conduct and of the market in which it is undertaken are such that it belongs to a class of conducts that distort the competitive process by *disadvantaging rivals* (i.e. through exclusionary effects, widely defined) or by enhancing market power (as in a concerted practice case) and, assuming a welfarist substantive standard, by establishing that the conditions present are such that a strong presumption can be made of adverse welfare effects. Alternatively, one can say that under a TEB LS the CA decides that there is liability by inferring adverse welfare effects from the potential of the conduct to distort the competitive process by disadvantaging rivals (i.e. through exclusionary effects, widely defined) or by enhancing market power (as in a concerted practice case).

Next we distinguish an *intermediate LS (ITFEB)* between the TEB LS and the *Full Effects Based (FEB)* LS under which liability is presumed after articulating a theory of harm to consumers (C.2), but without undertaking analysis of efficiency effects.

Finally the *Full Effects Based (FEB)* LS, represents the LS under which a finding of liability relies on all potential anticompetitive (exclusionary or market power enhancing) and also of the potential pro-competitive effects of the specific conduct being assessed and compared and a showing of adverse effects on *welfare* (consumer or total) of this specific conduct be then established. We distinguish two potential FEB LSs depending on whether the substantive standard is that of consumer or total welfare.

Table M2 below provides a summary of the various LSs.

Table M2: Identifying legal standards

³⁰ EU's by-object restrictions can be thought of as MPS LSs. These by-object restrictions differ from the US Strict Per Se LS in that they require the contextualization of the conduct (and that they are rebuttable) – see also below.

<i>Components of economic analysis applied in assessment</i>	<i>Set of Decisions</i>	<i>Legal Standards / Value of LSI indicator</i>
A	S1	Strict Per Se (SPS) LS: LSI = 1
A and B	S2	Modified Per Se (MPS) LS: LSI = 2
A and B and C.1	S3	Truncated Effects Based (TEB) LS: LSI = 3
A and B and C1 and C2	S4	LSI = 4, Intermediate between Truncated and Full Effects Based (FEB) LS (ITFEB) LSI = 5, FEB LS under a Consumer Welfare Substantive Standard LSI = 6, FEB LS under a Total Welfare Substantive Standard
A and B and C1 and C2 and C3	S5	
A and B and C1 and C2 and C3 and D	S6	

Source: Revised from Katsoulacos et.al. (2018)

3. Description of the datasets and conduct categories

3.1 The EC (DGCOMP) dataset

Antitrust decisions by the EC Competition Directorate (DGCOMP) for the period 1992 – 2016 were examined – that is, decisions under Art. 101 TFEU (on agreements) and 102 TFEU (on abuse of dominance). The EC dataset consists of 170 antitrust infringement decisions (excluding a very small number of decisions on exploitative abuses)³¹. While for much of the empirical analysis the whole dataset can be used, for constructing our LS indicators (LSI), following the methodology described in Section 2 above, a number of decisions (in which “gaps” in economic analysis are present – see also below), cannot be taken into account. As a result, the number of observations in our dataset drops to 160. 114 of the decisions included in our no-gaps in economic analysis dataset were appealed and challenged before courts.

3.2 The French dataset

The French Competition Authority (*Autorité de la concurrence*) is an administrative body created in 2009 out of the *Conseil de la Concurrence*'s transformation³². It acts in the name of the State, although not supervised by the Government in the exercise of its powers. It takes up cases either when referred to it by a plaintiff or ex-officio. According to the relevant reference to its website³³, the Authority's top priority remains consumer welfare. The FCA has the competence

³¹ For a full description of the EC dataset including non-infringement decisions, see Katsoulacos, Makri and Metsiou (2019).

³² We will refer to both of them as the “French CA”.

³³ <http://www.autoritedelaconcurrence.fr/user/index.php>

to apply both national and European legislation (Articles 101 and 102 of the TFEU). The FCA is perceived as being among the best European CAs³⁴.

The total number of French antitrust infringement decisions (excluding exploitative abuses) in our dataset is 282, from 2000 - 2017. Again, for constructing our LS indicators following the methodology described in Section 2, a number of decisions (in which gaps in economic analysis are present), cannot be taken into account. As a result, the number of observations in our dataset drops to 251. 144 of the decisions included in our no-gaps in economic analysis dataset were appealed and challenged before courts.

3.3 The Greek dataset

The Hellenic Competition Commission (“HCC”) is the independent authority responsible for the enforcement of Greek Law 3959/2011³⁵, "Protection of Free Competition" (“Competition Act”), previously Law 703/1977, as well as of Articles 101 and 102 of the Treaty for the Functioning of the European Union (TFEU). It takes up cases either when referred to it by a plaintiff or ex-officio. According to the Explanatory Note by the Ministry accompanying Law 703/1977, its purpose was to protect free competition in the marketplace for the benefit of the economy in general and of consumers in particular³⁶.

The Greek dataset consists of 77 antitrust infringement decisions (excluding decisions on exploitative practices), from 1997 - 2017. 69 of these were appealed and challenged before courts. The number of decisions in which no-gaps in economic analysis are present is 69, of which 61 were appealed and challenged before courts.

3.4 The Russian dataset

The Russian dataset is based on decisions following the judicial review of the Russian CA’s (the Federal Antimonopoly Service, FAS) antitrust infringement cases for the period 2008-2015. The first year of observation is the year following the new law ‘On protection of competition’ (that was better harmonized with European rules than the previous version was). All observations are claims to appeal infringement decisions made under art. 10 and 11 (including 11.1)³⁷ of the law. We use the content of judicial decisions instead of the FAS dataset. The main reason for this is that in Russia, commercial courts present the data on judicial review for all instances³⁸ in a very systematic and user-friendly manner. And, the dataset of courts’ decisions covers all of the decisions made by commercial courts. For the dataset of FAS decisions, this is not a case. This is a serious limitation given that the judicial review outcome forms an important part of our analysis.

³⁴ There were 187 employees (working on the competition enforcement) of the FCA in 2016, of which 52,4% lawyers, 21,4% economists, and 26,2% other experts.

³⁵ Year 2011 marked the abolition of Law 703/1977 (with consecutive amendments).

³⁶ There are no exclusions or exemptions from the general competition law (Law 3959/2011 which replaced Law 703/1977). The only existing sectoral exclusion concerns the telecoms sector. EETT (Hellenic Telecommunications and Post Commission) is entrusted with the competences to act as the Competition Authority in the electronic communications market (fixed and mobile telephony, wireless communications and Internet access providers) and the postal services market (postal and courier service providers).

³⁷ That concern, respectively, agreements including concerted practices and abuse of dominance practices.

³⁸ In Russia, there are four instances in commercial courts: first instance, appellate instance, cassation instance and highest (Supreme) court. Supreme Court has the discretion to consider or not consider cases. Supreme and cassation courts can send cases back to the first instance court. As a result, the number of instances regarding a case can be large (the maximum number in our dataset is 10).

The total number of decisions for all antitrust cases (excluding exploitative abuses) in the Russian dataset is 987. The number in which no-gaps in economic analysis are present is 539.

Conduct categories (or groups)

In the analysis below, we classify conducts into four conduct groups. These are:

- conduct group G1: this consists of hard-core restrictions which have strong market power-enhancing effects³⁹. They include price fixing, bid rigging, boycotts, market sharing and exclusive territories;
- conduct group G2: includes concerted practices and coordination;
- conduct group G3: consists of vertical restraints that, according to conventional wisdom, can have both competition restriction and welfare-enhancing effects; such as resale price maintenance (RPMmin, RPMmax), vertical exclusionary agreements, exclusive territories;
- conduct group G4: consists of exclusive contracts, tying and non-price discrimination, by dominant companies. Again, most of these can have, according to conventional wisdom, both competition restriction and welfare-enhancing effects.

The following Table 1, provides a description of the dataset for each country. Then, Table 2 describes the number of decisions in these countries depending on what is the highest level of economic analysis that is present in these decisions, from analysis level A to level D (Table M1). And, it gives the percentage of decisions with gaps in economic analysis in each of these decision sets. Some more comments on the decisions with/without gaps are in order.

As mentioned in the methodological section above, in order for the value of the LS indicator to reflect unambiguously the extent, and the type, of economic analysis that is utilized we must concentrate on decisions without gaps in economic analysis. Specifically, for constructing legal standard indicators (LSIs), the antitrust decisions in our samples are allocated into (6) sets, $S_i, i = 1, \dots, 6$, each of which contains only decisions assessed under a specific LS, from the strict Per Se (*set S₁*) to full effects-based (*set S₆*) – as shown in Table M2 above. Thus, the decisions in each consecutive set contain the economic analysis of the previous set plus one more additional step of economic analysis – these steps (or components of economic analysis) being described in Table M1. This implies that each set contains only decisions in which there are no gaps (or, no missing steps) in economic analysis. Of course, in our sample there are also decisions in which there are gaps (or, missing steps) in economic analysis. Gaps in economic analysis indicate that the CA puts forward arguments or claims to have reached a conclusion without having undertaken the analysis required in order to support this conclusion (e.g. putting forward arguments that consumers are harmed without having established first market-power enhancing or exclusionary effects). It is the sub-sample of *decisions without gaps* that we will be using for the construction of and in our analysis of LSIs below, though for the analysis of the types of economic analysis used in assessing various conduct groups and of their evolution we use the samples with the total number of decisions.

Table 1. Infringement, appealed and without gaps decisions per conduct group (totals and shares), for each country – shares in parentheses

	EC 1992-2016	Greece 1997-2017	France 2000-2017	Russia 2008-2015	Total

³⁹ In the EU they are considered as violations of the EU competition law art.101.

Number (share) of infringement decisions per conduct group NOTE: for Russia our observations are the appealed infringement decisions	Total	170	77	282	987	1516
	G1	99 (0,58)	18 (0,23)	84 (0,30)	352 (0,36)	553
	G2	6 (0,04)	13 (0,17)	91 (0,32)	157 (0,16)	268
	G3	32 (0,19)	25 (0,33)	41 (0,15)	261 (0,26)	358
	G4	33 (0,19)	21 (0,27)	66 (0,23)	217 (0,22)	337
Number (share) of appealed decisions per conduct group	Total	129 ⁴⁰ (0,76)	69 (0,90)	161 (0,57)	987 (1,00)	1346
	G1	85 (0,86)	16 (0,89)	47 (0,56)	352 (1,00)	500
	G2	3 (0,43)	11 (0,85)	50 (0,55)	157 (1,00)	221
	G3	20 (0,65)	24 (0,96)	25 (0,61)	261 (1,00)	330
	G4	21 (0,64)	18 (0,86)	39 (0,59)	217 (1,00)	295
Number (share) of decisions without gaps per conduct group ⁴¹	Total	160	69	251	539	1019
	G1	96 (0,60)	12 (0,17)	69 (0,27)	122 (0,23)	299
	G2	6 (0,04)	11 (0,16)	78 (0,31)	132 (0,24)	227
	G3	30 (0,19)	25 (0,36)	39 (0,16)	144 (0,27)	238
	G4	28 (0,18)	21 (0,30)	65 (0,26)	141(0,26)	255
Number (share) of appealed decisions without gaps (decisions with final court judgements) per conduct group	Total	114 (0,71)	61 (0,88)	144 (0,57)	539 (1,00)	858
	G1	77 (0,80)	10 (0,83)	41 (0,59)	122 (1,00)	250
	G2	2 (0,33)	9 (0,82)	41 (0,53)	132 (1,00)	184
	G3	18 (0,60)	24 (0,96)	24 (0,62)	144 (1,00)	210
	G4	17 (0,61)	18 (0,86)	38 (0,58)	141 (1,00)	214

Table 2. ‘Gaps’ in economic analysis in sets of decisions with successively increasing levels of economic analysis

Competition authority		The highest level of economic analysis present in the decision						Total
		A	B	C1	C2	C3	D	
EC	Total number of decisions	6	96	41	16	1	10	170
	Number of decisions without gaps	6	96	40	13	1	4	160
	% of decisions with gaps	0%	0%	2,4%	18,8%	0%	60%	5,9%
Greece	Total number of decisions	0%	16	34	25	2	0	77
	Number of decisions without gaps	0	16	34	17	2	0	69
	% of decisions with gaps	-	0%	0%	32%	0%	-	10%
France	Total number of decisions	39	29	120	82	12	0	282
	Number of decisions without gaps	39	29	109	63	11	0	251

⁴⁰ 9 of these decisions were pending when this analysis was undertaken (5 from G1, 1 from G2, 1 from G3 and 2 from G4).

⁴¹ The share indicated is the share in each conduct group in total decisions without gaps.

	% of decisions with gaps	0%	0%	9%	23%	8%	-	11%
Russia	Total number of decisions	204	130	543	108	2	0	987
	Number of decisions without gaps	204	130	171	34	0	0	539
	% of decisions with gaps	0%	0%	69%	69%	100%	-	45%

Source: calculated by authors using dataset

We notice (Table 1) that there are significant differences in the percentages of infringement decisions, associated with different conduct groups, between the countries. In EC there is a very clear predominance of G1 decisions associated with hard-core horizontal agreements (58%), while in other countries the enforcement effort is more evenly distributed to the 4 conduct groups, with G1 been about one third of decisions in France and Russia and only just 23% in Greece. On the other hand, the fraction of decisions in concerted practices G2 in EC is negligible (just 4%), while it is most significant in France, with a fraction (32%) which is about twice higher than in Greece and Russia. Vertical restraints (G3) decisions are most significant in Greece (33%) and Russia (26%), while they are less significant but still quite important in EC (18%) and France (15%). Abuse of dominance decisions are most significant in Greece (27%) and equally significant in France (23%) and Russia (22%), while in the EC it accounts for 19% of the decisions.

We also observe from Table 1 that there is a very high rate of appeals, that reaches a peak in Greece (90%), with 76% of the EC's decisions also appealed. The rate is much lower though still significant in France (57%). The distribution of appeals of decisions among the various conduct groups is quite even in Greece and France, but this is not true for the EC in which 86% of G1 decisions are appealed but the appeal rate is significantly lower for G2 (43%) and for G3 / G4 (about 65%).

The second part of Table 1 concerns statistics of infringement decisions and of appealed decisions for which there were no gaps (or no missing steps) in economic analysis. As noted above, this information is essential for building up statistics on the LS indicators and on annulment rates below. This part of Table 1 is best considered in conjunction with Table 2. One very interesting observation that we can make is the *significant difference between countries in the fraction of decisions with gaps in economic analysis*. The fraction of decisions in which gaps are present is very small in the EC (only 5,9%). The picture that emerges is that of a very coherent approach to assessment by the EC, building up its economic arguments by always incorporating all the necessary prerequisite elements of economic analysis in order to support its contentions. The fraction increases significantly in the case of Greece and France (10% and 11% respectively) and reaches extremely high levels in the case of Russia (45%). This finding provides a preliminary measure of the "enforcement quality" by the different CAs, indicating that it is high in the EC and low in Russia. A more satisfactory indicator of enforcement quality (which, however reinforces this conclusion) is a measure of how far are the LSs adopted by the CAs from their theoretically optimal levels. Below we present such measures of the quality of enforcement.

4. Economic analysis and legal standards in antitrust enforcement: empirical findings

4.1 The role of and the type of economic analysis applied in antitrust decisions

Table 3 below describes the role and type of economic analysis, i.e. the extent to which analysis components A to D shown in Table M1, are applied in decisions by each CA in each of the conduct groups. We show the total number of decisions in which an analysis component is used, in each conduct group and, in parenthesis, the share of the decisions in the total number of decisions in that conduct group. The sample used is that of the total decisions. It is important to note, concerning analysis components C3 and D, which concern analysis about potential efficiency effects of the conduct, that we do not take into account the efficiency defense under the EU competition law on agreements of art. 101(3)⁴².

Table 3. The role of and the type of economic analysis applied per conduct group, by jurisdiction (all decisions)

		Analysis components applied in reaching decisions					
		A=1	B=1	C1=1	C2=1	C3=1	D=1
G1	EC	99 (100%)	94 (94,9%)	15 (15,2%)	7 (7,1%)	0 (0,0%)	1 (1,0%)
	Greece	18 (100%)	18 (100%)	2 (11,1%)	8 (44,4%)	1 (5,6%)	0 (0,0%)
	France	84 (100%)	42 (50,0%)	14 (16,7%)	25 (29,8%)	1 (1,2%)	0 (0,0%)
	Russia	352 (100%)	64 (18,2%)	275 (78,1%)	53 (15,1%)	0 (0,0%)	0 (0,0%)
G2	EC	6 (100%)	6 (100%)	1 (16,7%)	0 (0,0%)	0 (0,0%)	0 (0,0%)
	Greece	13 (100%)	13 (100%)	8 (61,5%)	5 (38,5%)	0 (0,0%)	0 (0,0%)
	France	91 (100%)	77 (84,6%)	81 (89,0%)	32 (35,2%)	5 (5,5%)	0 (0,0%)
	Russia	157 (100%)	111 (70,7%)	99 (63,1%)	4 (2,5%)	0 (0,0%)	0 (0,0%)
G3	EC	32 (100%)	29 (90,6%)	20 (62,5)	3 (9,4%)	1 (3,1%)	1 (3,1%)
	Greece	25 (100%)	25 (100%)	22 (88,0%)	5 (20,0%)	0 (0,0%)	0 (0,0%)
	France	41 (100%)	38 (92,7%)	40 (97,6%)	18 (43,9%)	2 (4,9%)	0 (0,0%)
	Russia	261 (100%)	42 (16,1%)	146 (55,9%)	26 (10,0%)	1 (0,4%)	0 (0,0%)
G4	EC	33 (100%)	33 (100%)	30 (90,9%)	14 (42,4%)	4 (12,1%)	8 (24,2%)
	Greece	21	21	21	9	1	0

⁴² The rationale for this is that the burden of proof of efficiency benefits under art.101(3) falls on the defendants. For an analysis, concentrating on the sample of EC decisions, that takes into account of efficiency defense under art.101(3), see Katsoulacos and Makri (2020) – however, this makes difficult the comparison with non-EU countries. An important aspect of the empirical analysis of this paper is the comparison of EU countries to a younger non-EU jurisdiction (Russia). We can say that the focus of the present paper is to empirically identify the extent and type of economic analysis utilised by CAs to satisfy their burden of proof in their infringement decisions.

		Analysis components applied in reaching decisions					
		A=1	B=1	C1=1	C2=1	C3=1	D=1
		(100%)	(100%)	(100%)	(42,9%)	(4,8%)	(0,0%)
	France	66 (100%)	65 (98,5%)	65 (98,5%)	18 (27,3%)	4 (6,1%)	0 (0,0%)
	Russia	217 (100%)	125 (57,6%)	122 (56,2%)	25 (11,5%)	1 (0,5%)	0 (0,0%)
Total	EC	170 (100%)	162 (95,3%)	66 (38,8%)	24 (14,1%)	5 (2,9%)	10 (5,9%)
	Greece	77 (100%)	77 (100%)	53 (68,8%)	27 (35,1%)	2 (2,6%)	0 (0,0%)
	France	282 (100%)	222 (78,7%)	200 (70,9%)	93 (33,0%)	12 (4,3%)	0 (0,0%)
	Russia	987 (100%)	342 (34,7%)	642 (65%)	108 (10,9%)	2 (0,2%)	0 (0,0%)

Source: Calculated by authors using dataset

Apart from characterizing the conduct (analysis component A), which is present in all the decisions of all countries, a contextual market analysis (B) also characterises essentially all decisions (95,3%) of the EC. This is consistent with the convention (to be discussed again below) that in the EU even in by-object restrictions the CA must contextualise the conduct taking into account the situation in the market(s) in which it is undertaken⁴³. So the use of B is independent of the conduct group and the LS applied in that group. This is confirmed in Greece (100%), while in France the fraction is very high (about 78%) but still quite smaller than in the EC and Greece⁴⁴. In contrast, in only a relatively very small fraction (34,7%) of the decisions by the Russian CA there is contextual market analysis (B).

Table 3 gives rise to a number of very interesting observations related to the use of the other economic analysis components (C to D). Whether or not an economic analysis component C to D should be used in decisions concerning a specific conduct group depends on the LS that should be applied in the assessment of that conduct group. For a conduct group in which a Per Se LS (or by-object restriction) should be applied, analysis components C to D are not needed, while for a conduct group in which a full effects-based LS should be applied analysis components C to D are all important. When an intermediate LS is appropriate then some of the analysis components C should be present in the assessment but not necessarily all. This implies that one must interpret very carefully the fraction in the *total* decisions of a jurisdictions in which an analysis component is used (last 4 rows of Table 3), since what this fraction *should* be depends on the composition of the decisions across the conduct groups in each jurisdiction⁴⁵. A jurisdiction in which a very large fraction of decisions involve G1 cases should have a small fraction of decisions in which analysis components C to D are applied since the right LS for G1 cases is that of Per Se illegality. In contrast, in a jurisdiction in which a very large fraction of decisions involves G4 cases should have

⁴³ This, as already noted, been an important difference between this and the Strict Per Se illegality LS in US.

⁴⁴ The difference is likely to be mainly due to the fact that published decisions by the French CA tend to be much more economical in presenting the analysis actually undertaken than is the case for the Greek CA and for the EC.

⁴⁵ For example, we notice that in the EC, only in a seemingly very small fraction of total decisions (14,1%) the analysis component C.2 is present (much lower than in Greece and France), but this is just due to the EC having a much larger fraction of decisions in the G1 conduct group (in which C.2 should not be present) than the other countries.

a large fraction of decisions in which at least analysis components C1 and C2 are applied since the right LS for G4 cases is that of full effects-based.

Let us then turn to discuss the application of each one of the analysis components C to D.

- Analysis showing a market power enhancing or exclusionary effect (C1). We note that this is present in only 38,8% of the *total* number of decisions in the EC. However, this small fraction is due to the relatively very large fraction of G1 cases (hard core horizontal agreements) in EC, in which, rightly, no such analysis needs to be used. For the EC, the fraction of C1 is significant for conduct group G3 (62,5%) and particularly significant, as it should be, for conduct group G4 (90,9%) – conduct groups for which full effects-based should be used. C1 is present in a much larger fraction of *total* decisions (about 70%) in Greece and France reflecting the much lower fraction of G1 cases in these jurisdictions relative to the EC (Table 1). In these jurisdictions too (as in the EC) the fraction of C1 in decisions in the conduct groups G1, G3 and G4 is as it should be: very small in G1 (about 11% - about 17%) and very large in G3 and G4 which should be assessed by full effects-based (indeed for both G3 and G4 the fraction of C1 is higher in Greece and France than for the EC). On the other hand, in Russia, in a huge fraction of 78,1% of G1 cases C1 is used, when this should not be the case for a Per Se illegal conduct group, and a smaller fraction of C1 (of about 56%) in G3 and G4 cases, in which C1 should be applied. This already indicates that in this younger jurisdiction, the quality of enforcement has been very low. Finally, in the G2 conduct group, the fraction of C1 varies considerably across jurisdictions, from a low of 16,7% in the EC to a high of 89% in France. This is justified by the fact that, for the case of G2, the appropriateness of analysis component C1 depends on the exact nature of the cases examined. As we explain below, specifically for the EC the low fraction of C1 can be justified by the nature of the cases, been mostly by-object restrictions. Generally, however, G2 type cases should be assessed by effects-based, requiring analysis C1.
- Analysis showing consumer harm (C2). The performance of the EC here, for the overall period of time covered by our sample, is quite disappointing. While the EC rightly applies C2 in a very low fraction of decisions in groups G1 and G2⁴⁶, what is disappointing is that the fractions are small for G3 and for G4, where we would expect this to be an important feature of effects-based analysis: only in 9,4% of the G3 (vertical restraints) cases there is C2 analysis and for G4 the relevant fraction is just 42,4%. The picture for the EC, however, improves considerably when the evolution of the use of different types of economic analysis is examined – which shows a consistent and very substantial increase in C2 (associated with an increase in the LS adopted) for G4, over time⁴⁷. However, the picture does *not* show any significant improvement for the application of C2 in G3 decisions (vertical restraints) over time⁴⁸. In the other jurisdictions there is an unnecessarily high application of C2 in the G1 group category in Greece (44,4%) and France (29,8%), while it is relatively low in G3, where the fraction *should* be very high (20% for Greece and 43,9% in France) and in G4 (42,9% in Greece and 27,3% in France) – further, this low application of C2 in G4 does not increase over time, as it does in the case of

⁴⁶ Given the composition of its decisions in G2, which are mostly information exchange agreements involving future prices for which the right LS is Per Se illegality.

⁴⁷ See below for the increasing trend in the LS and, for more details for the evolution of each analysis component, Katsoulacos and Makri (2020).

⁴⁸ For more details see Katsoulacos and Makri (2020).

the EC⁴⁹. Russia performs worse, with C2 applied in only 10% of the cases in G3 and in 11,5% of the cases in G4 – without any tendency for these to increase over time⁵⁰.

- Efficiency effects analysis C3 (affecting consumers) and D (other efficiencies). This is also quite disappointing for the EC, though the examination of the evolution of the use of different types of economic analysis shows again quite a significant improvement for G4⁵¹. The presence of efficiency analysis is from negligible to very small: only in 12,1% of the G4 (abuse of dominance) cases there is such analysis for consumers (C3) and in 24,2% of the G4 cases there is any other efficiency analysis (D). The corresponding fractions for G3 are extremely low too: 3,1% (C3) and 3,1% (D). Overall, C3 is present in only 2,9% of the cases and D in only 5,9%. The situation in other jurisdictions is even worse with C3 appearing overall in 2,6% of the cases in Greece, 4,3% in France and 0,2% in Russia, with the percentage of D being 0% overall in all these other jurisdictions. Clearly, CAs still do not consider efficiency analysis an important component of their assessment of antitrust conducts, though the EC and France are doing relatively much better than Greece, with Russia performing very badly in this analysis component too.

4.2 The Weighted Average Legal Standards (WALS), their evolution and indicators of legal certainty and of the quality of enforcement

Empirical analysis in this section is based on the statistics presented in Tables 4, 5 and 6 below. To start with, in Table 4, we present the number (in parenthesis) and the share of the decisions (without gaps) that correspond to each value of LSI, $1 \leq LSI \leq 6$, for our four main conduct groups (G1 – G4). We also provide information (in square brackets) about the appealed decisions with final Court judgments, information that is used for the construction of Table 7 below.

Table 4. Number (in round brackets) and share of decisions without gaps that correspond to each LSI, by conduct group and jurisdiction. In square brackets the appealed decisions.

		LSI=1	LSI=2	LSI=3	LSI=4	LSI=5	LSI=6	Total no. of decisions
G1	EC	4,2% (4) [3]	82,3% (79) [66]	9,4% (9) [7]	4,2% (4) [1]	0,0% (0) [0]	0,0% (0) [0]	(96) [77]
	Greece	0,0% (0) [0]	83,3% (10) [8]	0,0% (0) [0]	8,3% (1) [1]	8,3% (1) [1]	0,0% (0) [0]	(12) [10]
	France	52,2% (36) [20]	29,0% (20) [11]	2,9% (2) [2]	14,5% (10) [7]	1,4% (1) [1]	0,0% (0) [0]	(69) [41]
	Russia	47,5% (58) [58]	15,6% (19) [19]	28,7% (35) [35]	8,2% (10) [10]	0,0% (0) [0]	0,0% (0) [0]	(122) [122]
G2	EC	0,0% (0) [0]	83,3% (5) [1]	16,7% (1) [1]	0,0% (0) [0]	0,0% (0) [0]	0,0% (0) [0]	(6) [2]

⁴⁹ See also Table 6 below for the evolution of the LS for G4.

⁵⁰ See also Table 6 below.

⁵¹ See Table 6 and, for more details, Katsoulacos and Makri (2020).

		LSI=1	LSI=2	LSI=3	LSI=4	LSI=5	LSI=6	Total no. of decisions
	Greece	0,0% (0) [0]	27,3% (3) [3]	45,5% (5) [3]	27,3% (3) [3]	0,0% (0) [0]	0,0% (0) [0]	(11) [9]
	France	2,6% (2) [1]	10,3% (8) [2]	50,0% (39) [24]	32,1% (25) [12]	5,1% (4) [2]	0,0% (0) [0]	(78) [41]
	Russia	15,9% (21) [21]	28,0% (37) [37]	56,1% (74) [74]	0,0% (0) [0]	0,0% (0) [0]	0,0% (0) [0]	(132) [132]
G3	EC	6,7% (2) [1]	33,3% (10) [6]	50,0% (15) [9]	6,7% (2) [1]	3,3% (1) [1]	0,0% (0) [0]	(30) [18]
	Greece	0,0% (0) [0]	12,0% (3) [3]	68,0% (17) [16]	20,0% (5) [5]	0,0% (0) [0]	0,0% (0) [0]	(25) [24]
	France	2,6% (1) [1]	0,0% (0) [0]	53,8% (21) [10]	38,5% (15) [11]	5,1% (2) [2]	0,0% (0) [0]	(39) [24]
	Russia	71,6% (103) [103]	8,3% (12) [12]	11,8% (17) [17]	8,3% (12) [12]	0,0% (0) [0]	0,0% (0) [0]	(144) [144]
G4	EC	0,0% (0) [0]	7,1% (2) [1]	53,6% (15) [9]	25,0% (7) [4]	0,0% (0) [0]	14,3% (4) [3]	(28) [17]
	Greece	0,0% (0) [0]	0,0% (0) [0]	57,1% (12) [9]	38,1% (8) [8]	4,8% (1) [1]	0,0% (0) [0]	(21) [18]
	France	0,0% (0) [0]	1,5% (1) [1]	72,3% (47) [31]	20,0% (13) [4]	6,2% (4) [2]	0,0% (0) [0]	(65) [38]
	Russia	15,6% (22) [22]	44,0% (62) [62]	31,9% (45) [45]	8,5% (12) [12]	0,0% (0) [0]	0,0% (0) [0]	(141) [141]
Total	EC	3,8% (6) [4]	60,0% (96) [74]	25,0% (40) [26]	8,1% (13) [6]	0,6% (1) [1]	2,5% (4) [3]	(160) [114]
	Greece	0,0% (0) [0]	23,2% (16) [14]	49,3% (34) [28]	24,6% (17) [17]	2,9% (2) [2]	0,0% (0) [0]	(69) [61]
	France	15,5% (39) [22]	11,6% (29) [14]	43,4% (109) [67]	25,1% (63) [34]	4,4% (11) [7]	0,0% (0) [0]	(251) [144]
	Russia	37,8% (204) [204]	24,1% (130) [130]	31,7% (171) [171]	6,3% (34) [34]	0,0% (0) [0]	0,0% (0) [0]	(539) [539]

Source: calculated by authors using dataset

A number of things stand out from a perusal of Table 4. As anticipated, and as should be the case, the LS for the vast majority of G1 decisions (over 80%) is that of Per Se or MPS (LSI = 1 or LSI = 2, respectively), for the EC, Greece and France⁵². In Russia, the fraction is quite low at about 62%, a result that confirms our finding above that in Russia in many G1 cases the CA applies, unnecessarily, analysis components above A (characterizing the conduct) and B (contextual analysis). For the G2 cases, a value of LSI = 2 for 83% of these cases in the EC is justified given the nature of G2 cases in the EC⁵³.

Then, a result that confirms the findings in Table 3 reported above, is the rather low LSs (closer to Per Se than to full effects-based, or $LSI \leq 3$) adopted, in the assessments of most G3 (vertical restraints) cases (90% in the EC, 80% in Greece, 54% in France). And, similarly, for G4 (abuse of dominance) cases (61% in the EC, 57% in Greece and 72% in France) - these two conduct groups been the ones for which, we would expect on the basis of the economic theory of optimal LSs that LSI would be on average higher (closer to effects-based). The situation is, as anticipated, even worse in Russia with about 92% of the cases in G3 and G4 been assessed with LSs corresponding to $LSI \leq 3$. However, another interpretation of the low LSs applied in the EU CAs is that because the period covered by our data for these CAs (1992 – 2016 for the EC) is long, and at least in the first decade of this period the effects-based approach did not have as strong theoretical support, and Per Se was still dominant, the averages shown in Table 4 conceal important differences in the LSs adopted over time. We present evidence below (Table 6) that this interpretation is certainly correct for group G4 for the EC for which there is a consistent and remarkable increase of the LS adopted over time, and to a certain though much lesser extent, for France, but not for Greece.

Another observation from Table 4 is that variations in the value of LSI are quite large. For example, for the EC, in group G3, in 50% of the cases $LSI = 3$ (i.e. the TEB LS is applied), while in another 33,3% of the cases $LSI = 2$ (i.e. a MPS LS is applied). In group G4, in 53,6% of the cases $LSI = 3$ (i.e. the TEB LS is applied), while in another 25% of the cases $LSI = 4$ (i.e. a ITFEB LS is applied). The variation in the value of LSI for G2 is significant for Greece and France and this is also true for G3 and for G4 for the EC, Greece and France. The variation in LSs adopted for each conduct group suggests that, from the point of view of firms, there is some legal uncertainty in the enforcement of competition law for these conduct groups, though not significant for these three jurisdictions given that in the vast number ($> 80\%$) of cases two neighbouring LSs are adopted in all conduct groups. Given that for these three jurisdictions the period examined is quite long, again, another interpretation of the variation in LSs for the same conduct group is that there is a trend towards higher (more effects-based) standards over time. This is certainly true for the EC for G4, and to some, though much smaller, extent for France (Table 6 below). In Russia, the absence of consistency in the use of legal standards in assessing any given group of conduct is larger and holds for all conduct groups, including G1. In Group G1, in 48% of the cases $LSI = 1$,

⁵² As we noted above, in contrast to the US where a strict Per Se LS (under which $LSI = 1$) is used for G1 cases, object-based restrictions in the EU require the CA to undertake contextual analysis (B) and hence, in our terms, expected LSI for G1 cases is $LSI = 2$ (which corresponds to what we have termed a MPS LS). We thank Prof. Pablo Ibanez Colomo for clarifying this point in discussions.

⁵³ As noted above, these are information exchange cases involving future prices. In other G2 cases, the more appropriate LS is effects-based (see for further discussion, Katsoulacos and Makri, 2020), which justifies the higher LSI for G2 cases in the other jurisdiction in Table 4 – though, for effects-based assessment these LSI values are too low for Russia (100% of cases $LSI \leq 3$) and closest to effects-based in France (in 87% of the cases $LSI \geq 3$).

while in another 29% of the cases $LSI = 3$. Further this is not due to an evolution towards higher standards in the period 2008 – 2015 (as shown in Table 6). This is another indication that the quality of enforcement is still very low in this jurisdiction. We return to the issue of legal uncertainty in Table 5.

Turning to Table 5, we present estimates of the weighted average legal standard (WALS) adopted as well as measures of both the quality of enforcement and of the extent to which there is consistency in the use of LSs in specific conduct groups. The WALS is calculated using Table 4: it is the sum of the values of LSI, each value weighted by its respective share in a given conduct group⁵⁴. The higher the WALS, the closer is the LS to effects-based (full effects-based, when the SS is total welfare, requiring a value of 6). With a Total Welfare (TW) substantive standard, $1 \leq WALS \leq 6$, while with a Consumer Surplus substantive standard $1 \leq WALS \leq 5$. The evolutions of WALS for each conduct group and jurisdiction is shown in detail in Table 6.

Since we will have to compare WALS to the optimal LS for any given conduct group we need to define the latter, taking into account our discussion of the literature on optimal LSs in the Introduction and the discussion above⁵⁵. Specifically, we will assume that the optimal LS for G1 is not strict Per Se (as in North America) but a restriction by-object (as in the EU), implying that the optimal LS corresponds to $LSI = 2$ (rather than $LSI = 1$). The optimal LS for G3 and G4 is full effects-based, so the optimal LS is given by $LSI = 6$ when the substantive standard is total welfare and by $LSI = 5$ when the substantive standard is consumer surplus. Concerning G2 (concerted practices), there are some conducts such information exchange agreements involving future prices for which there is more or less unanimity that they should be treated as object-restrictions (so here the optimal LS is given by $LSI = 2$), but for the other information exchanges⁵⁶ and concerted practices the optimal LS we will assume to be full effects-based. In the case of the EC all decisions involving G2 conducts concerned information exchange agreements and most of these involved future prices – so on average the optimal LS for G2 would be given by a value of $LSI = 3$ ⁵⁷. We assume that this is the relevant optimal benchmark for G2 in the other countries too.

The first observation is that, for the conduct group (G1) that is traditionally illegal Per Se (price fixing and market sharing), the value of WALS is, as expected, low. For the EC, Greece and France, but also for Russia, its value is closer to 2: for the EU CAs, this is consistent with how EU Competition Law interprets the appropriate use of economic analysis in by-object restrictions – as noted above, according to this interpretation, contextualizing the conduct (economic analysis B) is considered important even for by-object restrictions⁵⁸.

The second observation is that cases in group G2 (concerted practices) are treated in the EC with a (low) MPS LS (with $WALS = 2,17$), with the WALS for G2 been slightly higher in the other jurisdictions, reaching its highest value of 3,27 in France (still well below full effects-based). We assumed above that the value of the LSI that would be closest to the optimal for G2 cases is $LSI = 3$, though this depends on the composition of the cases in this conduct category, $LSI = 3$

⁵⁴ For example the value of WALS for G1, for Russia, in Table 5 is 1,98. This is obtained, using Table 4 as: $0,48*1+0,16*2+0,29*3+0,08*4 = 1,98$.

⁵⁵ Specifically in footnote 52.

⁵⁶ See for details Katsoulacos and Makri (2020), including a brief literature review on optimal LSs for information exchange agreements and references to the relevant literature.

⁵⁷ See for details, Katsoulacos and Makri (2020).

⁵⁸ Analysis under B can also be taken into account by the EC in considering the implications of various aggravating and mitigating circumstances for setting fines – e.g. the market shares of the undertakings and the geographic scope of the infringement.

been the optimal for the composition of the small number of the G2 cases in the EC (most involving exchange in information about future prices).

Thirdly, and contrary to theoretical predictions, group G3 (vertical restraints) with a WALs = 2,67 in the EC, is treated with a LS on average between MPS and TEB (without that is, establishing harm to consumers and, in some cases, not even a theory of exclusionary effects). The situation is broadly similar in Greece (WALS = 3,08) and France (WALS = 3,44). And it is much worse in Russia, for which WALs = 1,57. Further, it is important to note that in none of the jurisdictions examined there is evidence, in Table 6, of a systematic increase in the values of LSI for G3 conducts over time.

Finally, for cases in conduct group G4 (abuse of dominance) which should be treated, as G3, under full effects-based, the LS adopted is significantly lower than this social optimal benchmark, on average over the periods examined being 3,61 in the EC, 3,48 in Greece and 3,31 in France. In Table 6 we see that the low average in the EC is due to the very low value of LSI in early years (something that can be easily justified in the '90s decade) and that the LSI is very close to the optimal value for the EC in more recent periods. There is also some similar, though much weaker, evidence for this in France but certainly not for Greece. In Russia the WALs = 2,33 for G4 and there is not an increase worth noting over time (Table 6).

Table 5. The WALs and indicators of quality of enforcement and of legal certainty by conduct group and jurisdiction

		WALS	LSI with highest share	Two neighboring LSI with highest shares	Index of legal uncertainty	Concentration index (max. 1)	Quality (Q) of enforcement with TW SS, $0 \leq Q \leq 5$ (value of Q relative to max.deviation = 5)	Number of decisions
EC	G1	2,14	2	2&3	0,54	0,69	4,86 (0,97)	96
	G2	2,17	2	2&3	0,41	0,72	4,17 (0,83)	6
	G3	2,67	3	2&3	0,84	0,37	1,67 (0,33)	30
	G4	3,61	3	3&4	1,13	0,38	2,61 (0,52)	28
			Weighted Average Enforcement Quality (WAEQ) (relative to max.): 3,84 (0,77)					
Greece	G1	2,42	2	1&2 (2&3)	1,00	0,71	4,58 (0,92)	12
	G2	3,00	3	2&3 (3&4)	0,77	0,36	5,00 (100)	11
	G3	3,08	3	3&4	0,57	0,52	2,08 (0,42)	25
	G4	3,48	3	3&4	0,60	0,47	2,48 (0,50)	21
			WAEQ: 3,07 (0,61)					
France	G1	1,84	1	1&2	1,12	0,38	4,84 (0,97)	69
	G2	3,27	3	3&4	0,82	0,37	4,73 (0,95)	78

	G3	3,44	3	3&4	0,72	0,44	2,44 (0,49)	39
	G4	3,31	3	3&4	0,61	0,57	2,31 (0,46)	65
							WAEQ: 3,76 (0,75)	
Russia	G1	1,98	1	1&2	1,05	0,34	4,98 (0,996)	122
	G2	2,40	3	2&3	0,75	0,42	4,40 (0,88)	132
	G3	1,57	1	1&2	0,99	0,54	0,57 (0,11)	144
	G4	2,33	2	2&3	0,84	0,33	1,33 (0,27)	141
								WAEQ: 2,7 (0,54)

Source: calculated by authors using dataset

Two indicators in Table 5 measure the extent to which there is consistency in the use of LSs creating legal certainty. The first is the index of the concentration of legal standards (the HHI concentration index, calculated as the sum of the squared shares); so, the Concentration index is $HHI = \sum_i s_i^2$ where s_i – is a share of the particular legal standard i in the investigations of the particular conduct group, $\frac{1}{6} < \text{Concentration index} < 1$. The second is the standard deviation of LSI, an *index of legal uncertainty* calculated as the standard deviation of the LS, measured by LSI, applied in the assessment of a particular conduct group.

The concentration of LSI is rather low with no clear pattern as to its value across the conduct groups in the different jurisdictions. On average, concentration is highest, indicating highest consistency, in the EC (average of 0,54), next is Greece (with 0,52) and then France (0,44) and Russia (0,41). Thus for the Russian CA (FAS), the approach to assessment is least predictable. This ranking in the consistency in the use of LSs is confirmed with the index of legal uncertainty in which the lowest average score is that of the EC (0,73), followed by Greece (0,74), France (0,82) and in the last place is Russia (0,91).

In Table 5, two indicators measure the quality of enforcement. One (indicated by Q) is quality in assessing specific conduct groups. This is a measure of the divergence of WALS from the theoretically optimal LS. Since the optimal LS can be higher or lower than the standards actually used, Q is estimated as the difference between the maximum possible deviation and the (absolute value of the) actual deviation of WALS from the theoretically optimal LS. If the difference is equal to the maximum deviation (which implies a zero actual deviation) then Q is at its maximum. If the difference is zero (which implies that actual is equal to maximum deviation) then Q is at its minimum. We also express this difference relative to maximum deviation so its value lies between 0 (minimum quality) and 1 (maximum quality). Note that the maximum deviation depends on the substantive standard. With a total welfare substantive standard the maximum deviation is 5, since the maximum LS is 6 and the minimum actual LS is 1. With a consumer welfare substantive standard the maximum deviation is 4, since the maximum LS is 5. Thus, with a total welfare substantive standard Q ranges from $0 \leq Q \leq 5$ while with a consumer welfare substantive standard Q ranges from $0 \leq Q \leq 4$.

We also calculate an indicator of the *overall quality* of enforcement of competition law by the CA in the jurisdiction, across conduct groups. This is the Weighted Average Enforcement

Quality (WAEQ) index estimated by weighting the value of the quality indicator Q of each conduct group by the share of that group in the total number of decisions in the sample. With a total welfare standard, $0 \leq \text{WAEQ} \leq 5$ while with a consumer welfare standard, $0 \leq \text{WAEQ} \leq 4$.

We notice that in the EC the quality of enforcement is low for conduct group G3, for which quality Q, relative to the maximum deviation, is only 0,33. It is very high as would be expected, for G1 and for G2. For G4 the overall quality is low to medium (0,52) when the substantive standard is that of TW, though it can be easily shown that it is higher (0,65) when the substantive standard is that of CS. The relatively low value of Q for G4 in the EC, however, reflects more the use of very low LSs applied to G4 conducts in the earlier part of the period under consideration – as we can see when we examine the evolution of the WALs in Table 6 below that shows a systematic and remarkable increase in the value of WALs for G4 in the EC in more recent years towards the optimal level (indeed, reaching the optimal level of 5, under a CS substantive standard).

The value of Q is also very high in the other jurisdictions for G1 and G2 (been highest in Russia for G1 and highest in Greece for G2). For G3, the value of Q is generally low also (i.e. apart from been low in the EC) for Greece (0,42), France (0,49), and, extremely low, in Russia (0,11). Similarly the value of Q is low for G4 in France (0,46) and in Greece (0,5). Further, and importantly, since the value of WALs does not tend to converge to full effects-based for either G3 or G4 in more recent years, in both of these jurisdictions (Table 6), the value of Q also does not tend to increase – though it has to be recognized that the tendency towards effects-based is stronger in France, in which the value of WALs (for G3 and G4) is higher in recent years than the overall average.

Concerning, finally, the indicator of the overall quality of enforcement of the CA, WAEQ, this is moderately high to high in the EU (0,77) and France (0,75), while it is significantly (26%) lower in Greece (0,61) and even more so, at just WAEQ = 0,54, in Russia. Given the evolution of the value of WALs mentioned above, the superiority of the EC in the WAEQ is even stronger in more recent years.

Table 6. Evolution of the WALs by conduct group (number of decisions without gaps in parentheses)

	EC	Greece	France	Russia
G1				
1992-1995	2,21 (14)	-	-	-
1996-1999	2,00 (5)	-	-	-
2000-2003	2,04 (24)	-	2,00 (15)	-
2004-2007	2,00 (26)	2,67 (3)	1,74 (19)	-
2008-2011	2,19 (16)	2,75 (4)	1,76 (21)	1,82 (83)
2012-2015	2,55 (11)	2,00 (1)	2,00 (9)	2,31 (39)
2016-2017	⁵⁹	2,00 (4)	1,80 (5)	-

⁵⁹ One decision in 2016 has been added to the period up to 2015.

<i>Sample average</i>	2,14 (96)	2,42 (12)	1,84 (69)	1,98 (122)
G2				
1992-1995	2,50 (2)	-	-	-
1996-1999	2,00 (2)	-	-	-
2000-2003	-	2,33 (3)	3,12 (33)	-
2004-2007	-	3,40 (5)	3,33 (21)	-
2008-2011	2,00 (1)	-	3,23 (13)	2,43 (120)
2012-2015	2,00 (1)	-	3,71 (7)	2,17 (12)
2016-2017	-	3,00 (3)	3,50 (4)	-
<i>Sample average</i>	2,17 (6)	3,00 (11)	3,27 (78)	2,40 (132)
G3				
1992-1995	2,60 (10)	-	-	-
1996-1999	3,00 (6)	2,00 (1)	-	-
2000-2003	2,67 (9)	2,67 (3)	3,20 (10)	-
2004-2007	2,75 (4)	3,00 (6)	3,47 (17)	-
2008-2011	-	3,33 (9)	3,40 (5)	1,28 (86)
2012-2015	1,00 (1)	3,17 (6)	4,00 (4)	2,00 (58)
2016-2017	-	-	3,33 (3)	-
<i>Sample average</i>	2,67 (30)	3,08 (25)	3,44 (39)	1,57 (144)
G4				
1992-1995	2,75 (4)	-	-	-
1996-1999	3,00 (7)	-	-	-
2000-2003	3,43 (7)	3,50 (2)	3,53 (17)	-
2004-2007	4,40 (5)	3,67 (3)	3,16 (25)	-
2008-2011	4,33 (3)	3,55 (11)	3,00 (7)	2,17 (54)
2012-2015	5,00 (2)	3,20 (5)	3,30 (10)	2,44 (87)
2016-2017	-	-	3,67 (6)	-

<i>Sample average</i>	3,61 (28)	3,48 (21)	3,31 (65)	2,33 (141)
-----------------------	-----------	-----------	-----------	------------

4.3 LSs and the annulment rate of appealed decisions

As already mentioned, we are also interested on whether additional economic analysis, and hence higher LSs, are associated with higher annulment rates. That is, we are interested in testing the hypothesis that with effects-based LSs the *disputability* of the CA's decisions increases, which is expected to increase the probability of annulment – see, for example Neven (2006), who is probably the first to put forward this conjecture. This would be a factor influencing the choices of utility maximizing CAs, the reputation of which is affected adversely when their decisions are reversed by Appeal Courts (Katsoulacos, 2019a).

In Table 7 below, we present a number of annulment indicators. To start with the Table presents the annulment rate (AR) for each conduct group and each LSI:

$AR_{i,j}$ = annulment rate of decisions in conduct group j , $j=1,2,3,4$, for each LSI_i , $i=1,2,\dots,6$.

These are used to estimate and present the Weighted Average Annulment Rate (WAAR) for each of the LSI across the conduct groups for each jurisdiction. This is calculated as follows. Let:

$N_{i,j}$ = number of appealed decisions with final Court judgement in conduct groups j , $j=1,2,3,4$ for each LSI_i , $i=1,2,\dots,6$.

T_i = total number of appealed decisions with final Court judgement for each LSI_i , in all conduct groups. These are given in Table 4. $WAAR_i$ is then given by:

$WAAR_i = \sum_{j=1}^4 \frac{N_{i,j}AR_{i,j}}{T_i}$. The values of $WAAR_i$ for each LSI are shown, for each jurisdiction, in

Table 7 in the last four rows before the last row.

In Table 7 we also present the WAAR for each conduct group across all the LSI values,

$WAAR_j = \sum_{i=1}^6 \frac{N_{i,j}AR_{i,j}}{T_j}$ – i.e. this is given by the total number of annulled decisions in this conduct

group divided by the total number of appealed decisions in the conduct group. These are presented, for each jurisdiction, in the last column of Table 7 – where in parentheses we also show the WAAR across all jurisdictions. We also calculate and present in Table 7 the WAAR for each LSI for each conduct group across jurisdictions (in the blue highlighted rows) and in the last (yellow highlighted) row the WAAR for each LSI across all jurisdictions and across all conduct groups.

The findings, depicted in Table 7, do not provide, or provide very partial, corroboration of the hypothesis that when a more effects-based LS is adopted (as measured by an increase in the value of LSI), there is an increase in the AR (or WAAR) of the appealed decisions. Indeed, there is evidence that the opposite is true, certainly in the case of the EC. To see this we note the following:

- (i) Considering first each conduct type in isolation, the AR of appealed decisions do not show a clear pattern as LSI increases. For G1, the AR increases as LSI increases from LSI = 1 to LSI = 2 for the 3 jurisdictions for which there are decisions with LSI = 1 (EC, France and Russia). But then, the AR *declines* significantly when LSI increases from LSI = 2 to LSI = 3 for the EC decisions – for the other LSI values there are very few observations with 0 annulment rate. Though there are only 7 appealed decisions with LSI = 3, this indicates that the presence of economic analysis component C1, to show a market power raising effect, *lowered* the annulment rate. This is also true for Russia, with a significant reduction in the AR as LSI increases from LSI = 2 to LSI = 3. But then, for Russia, the AR shows a significant increase as the LSI increases further to LSI = 4. That is, there is a U-shaped

relation between LSI and the AR⁶⁰. On the other hand, in France there is a significant reduction in the AR as LSI increases from LSI = 2 to LSI = 3 and LSI = 4 (there is only one observation for higher LSI)⁶¹.

For G2, in the EC we have very few decisions and just 2 appeals without annulments. For Greece the AR first decreases and then increases though there is a very small number of 3 observations corresponding to LSI = 2,3,4. For France, the number of observations is also very small for LSI = 1,2 and then the annulment shows a considerable *increase* as LSI increases from LSI = 3 to LSI = 4 providing some confirmation for the Neven (2006) conjecture. On the other hand, for Russia the AR *decreases* significantly as LSI increases from LSI = 1, to LSI = 2, to LSI = 3.

For G3, in the EC, the AR *increases* considerably when LSI increases from LSI = 2 to LSI = 3 but it is zero for (the very small number) of decisions with higher LSI. For Greece on the other hand there is a continuous *decline* in the AR as LSI increases from LSI = 2 to LSI = 4, while in France there is a continuous *increase* as LSI increases from LSI = 3 to LSI = 5. For Russia, there is again a U-shaped relation as LSI increases from LSI = 1 to LSI = 3 but the annulment declines again for LSI = 4.

For G4, in the EC, the AR increases when LSI increases from LSI = 3 to LSI = 4 but it is then reduced considerably for LSI = 6 (*inverse* U-shaped relation). In Greece, neglecting the single decision with LSI = 5, the relation is more or less constant and similarly for France. For Russia, for conduct group G4 there is a strong U-shaped relation. Overall, for Russia, for conduct groups G1, G3 and G4, for which there are also decisions for which LSI is greater than LSI = 3, the evidence suggests a U-shaped relation.

(ii) Considering next the WAAR for each value of LSI across all the conduct groups (which is given by the 4 rows before the last one in Tables 7), we see that, for the EC, after a zero annulment for the 4 G1 decisions for which LSI = 1 (Table 4), the WAAR increases to 47,3% for LSI = 2⁶² but then *as LSI is increasing from LSI = 2 to LSI = 6, WAAR is not increasing, indeed, it is continuously decreasing*⁶³, indicating that increasing the economic analysis content of decisions on average reduces the likelihood that these decisions will be annulled if appealed. For Greece the WAAR also declines continuously as LSI increases from LSI = 2 (there are no decisions with LSI = 1), to LSI = 4⁶⁴. For France on the other hand, whilst WAAR is inverse U-shaped between LSI = 1 and LSI = 3, for LSI = 4 and LSI = 5 there is a clear and significant increase in the WAAR. Finally, for Russia, the WAAR follows a clear U-shaped relation with the WAAR minimized when LSI = 2 (when the rate is 40%).

(iii) Considering, also, the WAAR for each conduct group (which is given by the last column in Table 7), we see that, for the EC, this WAAR is lower for conduct group G4 (35,3%) than it is for conduct groups G1 (46,8%) and G3 (44,4%), even though, as we have seen from Table 5, the WALs is highest for G4 (i.e. while the LS adopted for G4 cases is closer to effects-based than for any other conduct group, this conduct group has the lowest

⁶⁰ A U-shaped relation can be explained in the case of a reputation maximizing CA that chooses LSs weighting between the expected loss in reputation from a reversal of its decisions and the increase in its cost per decision when LSI increases (see, for example, Katsoulacos, 2019a).

⁶¹ For Greece, essentially all G1 decisions are concentrated on LSI = 2.

⁶² 74 decisions – Table 4.

⁶³ For the other 36 (= 26+6+1+3) decisions (Table 4), with LSI greater than LSI = 2.

⁶⁴ There is an increase for LSI = 5, but there are only two decisions with this LSI value.

WAAR)⁶⁵. The inverse relation between WALs and WAAR observed in the EC, is not confirmed in Greece for which the WAAR for G1 (30%) is substantially lower than it is the WAAR for G3 (41,7%) which has a higher WALs (from Table 5). The increase in WALs for G4, relative to G3, on the other hand, is associated with a significant decline in WAAR⁶⁶. In France, the WAAR increases as WALs increases for G1, G2 and G3 but for G4 there is a decrease in WAAR (to 21,1%) relative to G1 and G2 (22% and 34,1%) though G4 has a higher WALs⁶⁷. Finally, for Russia, G1 and G3, that have the lowest WALs, have higher WAAR (56,6% and 50%) than G2 and G4 that have higher WALs but lower WAAR (46,2% and 41,8%)⁶⁸.

- (iv) It is also interesting to compare the numbers in parentheses in the last column of Table 7, giving the WAAR for each conduct group *across* jurisdictions. We see that this WAAR is lowest for G4 (36,4%), though the *average WALs for G4 across jurisdictions* is (from Table 5) highest, with a value equal to 3,2, relative to average WALs of about 2,7 for G2 and G3 and of about 2,1 for G1 - indeed the lowest average WALs, that of G1, produces about the same WAAR (of 47%) as the average WALs of G3. This is another indication that the conjecture of Neven (2006) is not confirmed.
- (v) Finally, we can compare the WAAR for each LSI for each conduct group across jurisdictions (the blue highlighted rows in Table 7). For G1 there is no clear pattern as LSI increases – WAAR increasing from 48,1% to 51% when LSI increases from LSI = 1 to LSI = 2 but *then declines to 38,6* for LSI = 3 and to 36,8 for LSI = 4. For G2 on the other hand there is a U-shaped relation emerging with very high WAAR (of 72,7%) for LSI = 1 that first declines and then increases after LSI = 3. As for G1, there is no clear pattern for G3, but it is interesting to notice the decline in WAAR for LSI above LSI = 3. Similarly there is no clear pattern for G4. Also, the last (yellow highlighted) row in Table 7, showing the WAAR for each LSI across conduct groups *and* across jurisdictions, does not produce a clear pattern though it is interesting to note that the values of all LSI greater than LSI = 3 are associated with *lower* WAAR than the values of LSI smaller than LSI = 3.

To conclude this section, we note that our empirical findings on the relation of LSs to annulment rates do not question the point that “the scope for disagreement is greater when economic theory and evidence are important” (Neven, 2006; Katsoulacos, 2019a). As intuition suggests, this must be indeed a force that would, *ceteris paribus*, increase the annulment rate when adopting more effects-based LSs. Our findings do indicate, however, that the relationship is much more complex and so predictions cannot rely on this intuition alone. For EU jurisdictions other important factors that must be taken into account are that by-object restrictions (in G1) require contextualization of the conducts, to support that the presumption of illegality is sufficiently strong, and also by-object restricted conducts are rebuttable: both of these can be strong forces tending to increase the annulment rate of decisions on such conducts.⁶⁹

⁶⁵ Indeed for the EC, there is an increase in WALs as move from G1 to G3 and then to G4 which is associated with a decline in the WAAR for each of these conduct groups – an inverse relationship (for the EC there are no annulled G2 decisions).

⁶⁶ So, for Greece the relation appears to be inverse U-shaped.

⁶⁷ So the relationship appears to be inverse U-shaped in France too (as for Greece) when considering WAAR and WALs for the different conduct groups.

⁶⁸ More specifically, the relation between WALs and the WAAR for the different conduct groups in Russia is increasing, then decreasing and then increasing again (it is U-shaped for WALs exceeding the value of 1,98 for G1).

⁶⁹ For a more detailed discussion of this and an explanation of the inverse relation that seems to hold for the EC, see Katsoulacos and Makri (2020).

Table 7. Annulment rates and weighted average annulment rates (WAAR) for each particular LSI over the conduct groups, for each conduct group and across jurisdictions (number of decisions without gaps)

Conduct group	Jurisdiction	Indicator	LSI=1	LSI=2	LSI=3	LSI=4	LSI=5	LSI=6	WAAR for each conduct group across LSI, for each jurisdiction (across jurisdictions ⁷⁰)
G1	EC	Annulment rate for each LSI	0,000	0,515	0,286	0,000	-	-	0,468
	Greece	Annulment rate for each LSI	-	0,375	-	0,000	0,000	-	0,300
	France	Annulment rate for each LSI	0,200	0,364	0,000	0,000	1,000	-	0,220
	Russia	Annulment rate for each LSI	0,603	0,632	0,429	0,700	-	-	0,566
WAAR for each LSI for conduct group G1 across jurisdictions			0,481	0,510	0,386	0,368	0,500		(0,468)
G2	EC	Annulment rate for each LSI	-	0,000	0,000	-	-	-	0,000
	Greece	Annulment rate for each LSI	-	0,333	0,000	0,333	-	-	0,222
	France	Annulment rate for each LSI	1,000	0,500	0,250	0,417	0,500	-	0,341
	Russia	Annulment rate for each LSI	0,714	0,486	0,378	0,0	-	-	0,462
WAAR for each LSI for conduct group G2 across jurisdictions			0,727	0,465	0,333	0,400	0,500		(0,418)
G3	EC	Annulment rate for each LSI	0,000	0,167	0,778	0,000	0,000	-	0,444
	Greece	Annulment rate for each LSI	-	1,000	0,375	0,200	-	-	0,417
	France	Annulment rate for each LSI	1,000	-	0,300	0,364	0,500	-	0,375
	Russia	Annulment rate for each LSI	0,524	0,417	0,471	0,417	-	-	0,500
WAAR for each LSI for conduct group G3 across jurisdictions			0,524	0,429	0,462	0,345	0,333		(0,471)
G4	EC	Annulment rate for each LSI	-	0,000	0,333	0,500	-	0,333	0,353
	Greece	Annulment rate for each LSI	-	-	0,222	0,250	1,000	-	0,278
	France	Annulment rate for each LSI	-	0,000	0,226	0,250	0,000	-	0,211
	Russia	Annulment rate for each LSI	0,636	0,274	0,467	0,583	-	-	0,418
WAAR for each LSI for conduct group G4 across jurisdictions			0,636	0,266	0,351	0,429	0,333	0,333	(0,364)
WAAR for each LSI across conduct groups for EC			0,000	0,473	0,462	0,333	0,000	0,333	
WAAR for each LSI across conduct groups or Greece			-	0,500	0,286	0,235	0,500	-	
WAAR for each LSI across conduct groups for France			0,273	0,357	0,239	0,294	0,429	-	
WAAR for each LSI across conduct groups for Russia			0,578	0,400	0,421	0,559	-	-	
WAAR for each LSI across jurisdictions and across all conduct groups			0,539	0,427	0,370	0,385	0,400	0,333	

⁷⁰ In parentheses we show the WAAR for each conduct group across all the jurisdictions.

5. Conclusions

For the purposes of this paper we have collected and analysed a unique dataset of antitrust infringement decisions reached by four CAs in the last 25 years (EC's DGCOMP and the Greek, French and Russian CAs). We also traced the outcome, for all the decisions that were appealed, of the judicial review process – up to the final decision reached by the Appeal Courts. Our analysis contributes to the development of tools for the empirical measurement of the extent of economic analysis and legal standards in competition law enforcement. Our second contribution consists in a detailed empirical description of the role of economic analysis in competition law enforcement. Our approach allows us to precisely assess the role of economic analysis in comparison to judgments based on anecdotal evidence. We show that, on average, over the period analysed, economic analysis plays a relatively modest role in CA investigations, though with quite significant differences between jurisdictions, with LSs that are closer to *Per Se* (especially in Russia). Further, there is no discernible evolution towards full effects-based for conduct groups for which this would be the appropriate LS, with the important exception of EC for the abuse of dominance conduct group. Also, inconsistency in the choice of LSs for specific conducts can, in some cases, create legal uncertainty for firms, this been most serious in the case of Russia. Overall, our empirical findings indicate *varying levels of quality of enforcement in terms of the appropriateness of the LSs adopted, quality been considerably higher in the EC and France, than in Greece and in Greece than in Russia.*

Our investigation of the relation between LSs adopted and the outcome of the judicial review of decisions appealed, i.e. the rate of decision annulment, shows no consistent support for Neven's (2006) conjecture that moving towards effects-based increases this rate because “the scope for disagreement is greater when economic theory and evidence are important”. While in some cases, when LSs increase beyond a certain point towards effects-based the annulment rate increases, we certainly cannot verify this relation for the EC where, indeed, the opposite seems to hold, while for Russia there is evidence for most conduct groups of a U-shaped relationship.

References

- Avdasheva, S., Golovanova, S., & Katsoulacos, Y. (2018). The Impact of Performance Measurement on the Selection of Enforcement Targets by Competition Authorities: The Russian Experience in an International Context. *Public Performance & Management Review*, 1-28.
- Avdasheva, S., Golovanova, S., & Korneeva, D. (2016). Distorting effects of competition authority's performance measurement: the case of Russia. *International Journal of Public Sector Management*, 29(3), 288-306.
- Baker, J. B. (2003). The case for antitrust enforcement. *Journal of Economic Perspectives*, 17(4), 27-50.
- Beckner Frederick C. III & Steven C. Salop (1999). *Decision Theory and Antitrust Rules*, 67 *Antitrust L.J.* 41-76.
- Blair, R. D., & Sokol, D. D. (2012). The Rule of Reason and the Goals of Antitrust: An Economic Approach. *Antitrust LJ*, 78 (2), 471-504.
- Carlton, D. W. (2007). Does antitrust need to be modernized? *Journal of Economic Perspectives*, 21(3), 155-176.

- Easterbrook Frank H. (1984). "The Limits of Antitrust", 63 (1) TEX. L. REV. 1-40.
- Gavil, A. I. (2008). Burden of Proof in US Antitrust Law. *Issues in Competition Law and Policy*, ABA Section of Antitrust Law.1, 125-157.
- Geradin, D. & N. Petit. Judicial Review in European Union Competition Law: A Quantitative and Qualitative Assessment. TILEC Discussion Paper No. 2011-008; Tilburg Law School Research Paper No. 01/2011. Available at SSRN: <https://ssrn.com/abstract=1698342>
- Gerber, D. J. (2015). Adapting the role of economics in competition law: a developing country dilemma. In: *The Economic Characteristics of Developing Jurisdictions: Their Implications for Competition Law*, Ed. By M.S.Gal and others. pp.248-264.
- Gifford Daniel J. and Robert T. Kudrle (2015). The Atlantic Divide in Antitrust: An Examination of US and EU Competition Policy. *University of Chicago Press*.
- Jones, A., & Kovacic, W. E. (2017). Identifying Anticompetitive Agreements in the United States and the European Union: Developing a Coherent Antitrust Analytical Framework. *The Antitrust Bulletin*, 62(2), 254-293.
- Katsoulacos Y. (2019a). On the Choice of Legal Standards: a Positive Theory for Comparative analysis. *European Journal of Law and Economics*, 48(2), 125 - 165.
- Katsoulacos, Y. (2019b). A Note on the Concepts of Legal Standards and Substantive Standards (and how the latter influences the choice of the former). *Journal of Antitrust Enforcement*, 7(3), 365-385.
- Katsoulacos Y., Avdasheva S., Golovanova S. (2016). Legal standards and the role of economics in Competition Law enforcement. *European Competition Journal*, 12 (2-3), 277-297.
- Katsoulacos Y., Avdasheva S., Golovanova S. (2019). A methodology for empirically measuring the extent of economic analysis and evidence and identifying the legal standards in competition law enforcement. In: *Festschrift in Honor of Frederic Jenny, Concurrences Review*.
- Katsoulacos Y. and G. Makri (2020). The role of economics and the type of legal standards in antitrust enforcement by the EC: an empirical investigation. DP available in CRESSE Publications <http://www.cresse.info/default.aspx?articleID=3388>.
- Katsoulacos Y, G. Makri and E. Metsiou (2019). "Antitrust Enforcement in Europe in the Last 25 Years: Developments and Challenges", *Review of Industrial Organisation*.
- Katsoulacos, Y., Metsiou, E., & Ulph, D. (2016). Optimal substantive standards for competition authorities. *Journal of Industry, Competition and Trade*, 16(3), 273-295.
- Katsoulacos, Y., & Ulph, D. (2009). On optimal legal standards for competition policy: a general welfare based analysis. *The Journal of Industrial Economics*, 57(3), 410-437.
- Katsoulacos, Y., & Ulph, D. (2011). Optimal enforcement structures for competition policy: Implications of judicial reviews and of internal error correction mechanisms. *European Competition Journal*, 7(1), 71-88.
- Katsoulacos, Y., & Ulph, D. (2016). Legal uncertainty, competition law enforcement procedures and optimal penalties. *European Journal of Law and Economics*, 41(2), 255-282.
- Katsoulacos, Y., & Ulph, D. (2017). Regulatory decision errors, Legal Uncertainty and welfare: A general treatment. *International Journal of Industrial Organization*, 53 (2), 326-352.
- Lianos, I. (2012). The emergence of forensic economics in competition law: Foundations for a sociological analysis. University College London, Centre for Law, Economics and Society , Research Paper Series: 5/2012

- Neven D. J (2006). “Competition economics and antitrust in Europe”, *Economic Policy*, Volume 21 (48), 1 October 2006, Pages 742–791.
- Padilla, J. (2010). The elusive challenge of assessing information sharing among competitors under the competition laws. *Roundtable on Information Exchanges Between Competitors Under Competition Law*, 879-924.
- Rey P. and J.S. Venit (2015), An Effects-Based Approach to Article 102: A Response to Wouter Wils, *World Competition*, 38(1), 3-30.
- Shastitko, A. (2018). Empirical assessment of the role of economic analysis in the Russian antitrust: Why is economic analysis used? *European Journal of Law and Economics*, 45(2), 313-330.
- Schinkel, M. P. (2008). Forensic economics in competition law enforcement. *Journal of Competition Law and Economics* 4(1): 1-30.
- Sokol D. (2017). Troubled Waters between US and European Antitrust. Review of Gifford and Kudrle, *Michigan Law Review*, Vol. 115 (6), 955 - 977.
- Vickers, J. (2005). Abuse of market power. *The Economic Journal*, 115(504).
- Wils W. (2014), The Judgment of the EU General Court in Intel and the so-called More Economic Approach to Abuse of Dominance, *World Competition* 37 (4), 405-434.