COMISIÓN NACIONAL DE LOS MERCADOS Y LA COMPETENCIA

GUIDE FOR QUANTIFYING HARM FROM COMPETITION LAW INFRINGEMENTS

G-2020-03

Date: 11 July, 2023.

www.cnmc.es

CONTENTS

PRESENTATION

One of the objectives of the Spanish National Markets and Competition Commission (CNMC) is to guarantee effective competition in all sectors and markets and to disseminate a culture of competition that benefits consumers and users.

Anticompetitive business practices have a negative impact on the functioning of markets and the economy, undermining competitiveness, growth, innovation and job creation. Such behaviour can also cause specific harm to market participants (consumers and users, private operators, and public authorities), who may be deprived of the benefits they would obtain in the absence of anticompetitive behaviour.

Competition (or antitrust) law provides channels through which stakeholders who have suffered harm as a result of anticompetitive behaviour can seek compensation, either through private actions seeking a declaration of wrongdoing or via court proceedings to claim damages for breach of competition law infringements. These actions and claims can in turn play an important role in deterring anticompetitive behaviour and support the efforts of competition authorities.

However, the effectiveness of these initiatives, can be hampered by the complexity of determining the amount of harm suffered in each individual case.

Given this situation, the purpose of this Guide is to facilitate the task of quantifying the harm caused by competition law infringements for all parties involved: judges and courts, lawyers and experts specialising in this field, and actual or potential victims of anticompetitive conduct.

To achieve this objective, the Guide, which is purely advisory in nature, adopts a fundamentally economic and integrative approach. It presents the main economic, statistical and econometric concepts in a simple and instructive manner, without sacrificing rigour. It also aims to facilitate the task of preparing and analysing expert reports using practical examples and checklists.

The Guide is part of Strategic Line 10 of the CNMC's 2023 Action Plan.

1. INTRODUCTION

1.1. The role of CNMC in the private enforcement of competition law

1. **Competition law** is a **twofold** branch of law: **public**, when public administrative bodies such as the CNMC enforce competition law regulations in the public interest; and **private**, when injured parties go to the ordinary courts to seek compensation for possible harm related to competition law infringements. For a long time, in Europe, public enforcement of competition law has prevailed as a guarantor of the efficiency and proper functioning of national and European Union markets. It is within this scope of public enforcement that a large part of the CNMC's functions fall, as set out in Article 5 of Law 3/2013, of 4 June, on the creation of the National Commission for Markets and Competition (hereinafter, LCCNMC after its name in Spanish).

2. In recent years, and especially since the entry into force of EU Directive 2014/104/EU[1], better known as the **"Damages Directive"**, private enforcement of competition law has become increasingly important, as Article 3 of the Directive expressly recognises the right to full compensation for damages caused by infringements of competition law. In the same vein, both the EU Court of Justice and the Spanish courts have repeatedly ruled that any person who has suffered harm as a result of an infringement of competition law, i.e., of Articles 101 or 102 of the Treaty on the Functioning of the European Union (hereinafter, TFEU) or of Articles 1 and 2 of Law 15/2007 on the Defence of Competition (hereinafter, LDC), has the right to claim full compensation for the harm caused by the infringement[2]. The aim is to restore the injured party to the situation in which they would have been had there been no infringement.

3. **Full** compensation consists of **three components**: actual loss or *damnum emergens* (the decrease in wealth caused by the infringement); or *lucrum cessans* (the increase in wealth that would have occurred in the absence of the infringement); and the payment of interest (the capitalisation of the amount claimed as compensation for past harm at the time the harm is assessed). The purpose of quantifying harm is, therefore, to calculate the difference in the victim's wealth between the actual scenario (with

---

[1] Directive 2014/104/EU of the European Parliament and of the Council, of November 26, 2014, regarding certain rules governing actions for damages under national law, for infringements of the competition law of the Member States and of the European Union. The transposition of this Directive into Spanish law took place through the Royal Decree-Law 9/2017, of May 26.

[2] As established in Article 71.2 a) of the LDC.

infringement) and the counterfactual scenario (the situation without the infringement). The main challenge is to correctly define the counterfactual to be able to quantify the effect of the infringement.

4. In this context, **CNMC's role** in private enforcement of competition law is advisory in nature and is regulated in Article 15 bis of Law 1/2000 on Civil Proceedings (hereinafter, LEC, after its Spanish name) in its *amicus curiae*[3] facet and in Articles 5.2(b) of the LCCNMC and 76.4 of the LDC, introduced by the transposition of the Damages Directive, which empower the competent judicial bodies to request the CNMC to report on the criteria for quantifying harm[4].

5. It is therefore necessary to explain the most appropriate criteria for quantifying harm in the context of the anticompetitive practice. In this way, the CNMC plays an **advisory role** and does not participate in the judicial process as an interested party, but rather assists the court by providing information, experience and technical knowledge. This advisory role is not to be confused with the calculation, quantification or specific and individual assessment of the compensation, which is determined by the competent judicial body.

6. Finally, the CNMC considers that private enforcement of competition law is of the utmost importance in promoting **a culture of competition**, as it ensures that the positive effects of effective competition reach all economic agents, and deters operators from implementing anticompetitive practices. Therefore, as part of the tasks entrusted to the CNMC in the promotion of competition, it has deemed it appropriate to develop this Guide in accordance with the functions set out in Article 5.1, Section h) of the LCCNMC: *"To promote and carry out studies and research on competition matters, as well as general reports on economic sectors."*

---

[3] Pursuant to the provisions of Article 15.3 of Council Regulation (EC) No. 1/2003 of December 16, 2002, on the application of the rules on competition provided for in Articles 81 and 82 of the Treaty, Article 15 bis of the LEC allows the European Commission, the CNMC and the competent bodies of the Autonomous Communities to intervene, without being a party, on their own initiative or at the request of the judicial body, by providing information or submitting written observations on matters relating to the application of Articles 101 and 102 of the TFEU or Articles 1 and 2 of the LDC. With the permission of the corresponding judicial body, the competition authorities may also present verbal observations.

[4] In fact, the current Article 76.4 of the LDC is not completely new, since Article 25 c) of the LDC and Articles 25 h) and 13.3 of Law 16/1989 (after Law 52/1999) already included this advisory role of the Court for the Defence of Competition/National Competition Commission to rule on criteria for quantifying compensation.

1.2. Aim and target group of the Guide

7. The Guide has several **objectives**: to assist judges and courts quantify the compensation for harm, and to disseminate good practices to all stakeholders that take part in the proceedings that quantify harm in terms of competition law infringements. It should be emphasised that the content of this Guide is merely advisory (it is not legally binding) and not exhaustive, subject to future developments in the field.

8. **The Guide is therefore intended for a wide audience**: judges and courts dealing with antitrust damages **actions** arising from competition law infringements; the parties involved in these legal proceedings; economic experts and lawyers specialised in this type of claim; and the general public.

9. Currently, there are other guides and materials dealing with the quantification and estimation of harm. However, the CNMC considered it appropriate to produce this Guide, based on the consensus adopted in these and other references, to complement them and add value, in an attempt to facilitate the task of understanding and preparing expert reports. In this way, the added value of the Guide could be summarised as follows:

   i. The Guide brings together the **relevant guidelines** on how to quantify harm, which are scattered in different manuals and guides, making it difficult to access the relevant information[5].

   ii. The Guide includes a section on **expert reports**, which provides a series of recommendations related to their structure and content, so that they are as explanatory as possible. Far from increasing the obligations of the parties, this section is intended to convey best practice.

   iii. To facilitate the analysis of the robustness and consistency of the results of expert reports, the Guide includes a **checklist**, with a series of methodological precautions to be applied when using quantification methods to verify the reliability of the results.

   iv. To facilitate the understanding of more technical concepts, the Guide includes two annexes: the first contains a **glossary of economic**

[5] To illustrate this, the following can be consulted: European Commission (2013, 2019 and 2020), CNMC (2018 and 2020) and the Mercantile Court of Barcelona (2019). Other materials have also been taken as reference, such as OECD (2011) and the manuals prepared by Oxera (2009) and RBB Economics and Cuatrecasas (2017) at the request of the European Commission.

terms, while the second covers **statistical and econometric concepts**.

v. To illustrate the **practical application** of the most commonly used quantification methods (comparative methods), the Guide includes a third annex that **reviews the relevant economic literature** and a fourth that provides a very detailed model of a **practical example**.

1.3. Brief legal background on the compensation for harm

10. The adoption of the Damages Directive was an important milestone in this field and, although it has not fully harmonised damages claims procedures in terms of competition law infringements throughout the European Union, it has established guidelines that ensure greater procedural uniformity and efficacy in the Member States, all based on the principles of effectiveness and equivalence[6].

11. The **transposition of the Damages Directive** into Spanish law was articulated through Royal Decree-Law (RDL) 9/2017 (hereinafter, RDL 9/2017), in force since 27 May 2017. This RDL introduced changes to existing legislation, specifically to the LDC, in substantive or material matters[7], and to the LEC, in procedural aspects regarding access to sources of evidence[8]. However, it should be borne in mind that there are still cases that are currently governed by the framework that pre-dates the adoption of the Damages Directive.

12. Of particular relevance is Article 72 of the LDC, which recognises the **right to full compensation** for harm suffered in the event of an infringement of competition law. This compensation covers, from the perspective of material (pecuniary) harm, compensation for actual loss and loss of profits, as well as the payment of interest. However, the affected party may only claim the overcharge actually borne that has not been passed on and which has

---

[6] Regarding the "principle of effectiveness", Article 4 of the Directive establishes that the configuration of the rules and procedures governing damages claims in each Member State is configured in such a way that it is possible to bring an action. Regarding the "principle of equivalence", that same article establishes that in the event of a damages claim for infringement of European competition law, the injured parties must receive the same treatment that they would receive in the event of a damages claim derived from similar infringements but covered by domestic law. However, it should be borne in mind that both principles are included in the second additional provision of Royal Decree-Law 9/2017, which transposed the Damages Directive.

[7] Specifically, through the transposition of the Directive, Title VI was incorporated into the LDC under the name "Compensation for harm caused by practices restricting competition", together with a specific regulation on damages claims for infringements of competition law in Spain.

[8] Spanish Civil Law (LEC, after its Spanish name) introduced a new Section 1 bis ("Access to sources of evidence in damages claims procedures for infringement of competition law") within Chapter V, Title I, Book II.

caused harm, in accordance with the provisions of Article 78.1 of the LDC. This Guide focuses on quantifying **material** harm, which is that which is usually claimed for, without prejudice to the possibility of claiming other damages, where appropriate.

13. To facilitate the filing of a damages claim, Article 74.1 of the LDC establishes a **limitation period of five years**. Furthermore, Article 76.2 of the LDC **empowers the courts to estimate the amount** of the damages claim if it is proven that the affected party suffered harm, but it is practically impossible or excessively difficult to quantify them precisely on the basis of available evidence.

14. In addition, Article 75.1 of the LDC grants **"irrefutable" evidentiary value** to competition law infringements determined by a **final decision** of a Spanish competition authority[9] (including the CNMC) or a Spanish court for the purposes of the damages claim to be brought. It is important to note that Article 76.3 of the LDC provides for a rebuttable *iuris tantum* presumption that infringements qualified as cartels cause harm. Likewise, Article 75.2 establishes a rebuttable *iuris tantum* presumption concerning infringements established in a final decision of a competition authority or court of any other Member State, "without prejudice to the possibility of alleging and proving new facts of which it was not aware in the original proceedings".

15. It is also worth noting that the parties will have **access to sources of evidence** –in the possession of both the opposing party and the competition authority itself– so that they can better defend their interests and thus try to solve the problem of information asymmetry. The regulation is found in Articles 283 bis a) and following the LEC.

16. Finally, regarding **liability**, Article 73 of the LDC declares joint and several liability in the case of several infringers, excepting from the general rule those companies that are either small or medium-sized[10] and which meet certain requirements, or have been exempted from paying the fine. According to Article 73(4), parties benefiting from exemption from fines under a leniency programme are generally liable to their direct or indirect purchasers or suppliers and are only jointly and severally liable to other injured parties where full compensation cannot be obtained from the other undertakings involved in the same infringement of competition law.

---

[9] Damages claims based on a decision by a competition authority are known as follow-on claims. Otherwise, these are stand-alone damages claims.

[10] In accordance with the definition given in Commission Recommendation 2003/361/CE of May 6, 2003, on the definition of micro, small and medium-sized companies.

2. QUANTIFICATION OF HARM

2.1. General issues

2.1.1. Anticompetitive conduct and the agents involved.

17. In accordance with the provisions of the LDC (Article 71), there are two types of anticompetitive conduct; these are sanctioned both in the TFEU (Articles 101 and 102) and in Spanish law through the LDC (Articles 1 and 2) and may give rise to claim for damages. Such conduct consists of:

i. **Collusive practices:** collective agreements, concerted practices, decisions or recommendations for price **fixing, quantities, trading conditions or market sharing**[11]**.** A particularly serious example of this type of unlawful conduct is a cartel. In general, these practices disrupt the equilibrium in the market and lead to purchasers paying a price premium for products or services purchased compared to a situation without an infringement[12]. The potential price increase of products may also be accompanied by a reduction in the quantities sold[13].

ii. **Abuse of a dominant position**: conduct by one or more companies with a dominant position that restricts or impedes competition in the market. Two types of abuses are typically distinguished:

- **Exclusion abuse**: this has the effect of totally or partially excluding actual or potential competitors. It can take many forms, including predatory pricing, certain exclusionary discounts, tied selling, bundling, vertical exclusivity agreements, refusal to supply or margin squeezing.

- **Exploitative abuse**: this allows the dominant company to obtain unjustified advantages from its customers or suppliers by directly or indirectly imposing unfair prices or terms of trade. Examples include discriminatory practices and excessive prices.

---

[11] Depending on where the colluding companies are located within the production or distribution chain, a distinction is usually made between horizontal agreements (e.g., cartels or cooperation agreements) and vertical agreements (e.g., resale price fixing, single brand or exclusive distribution agreements).

[12] There is also the possibility that several operators will agree to lower purchasing prices to increase their profits and thereby harm their suppliers. For a practical example of quantifying harm in these cases, see Daggett and Freedman (1984).

[13] In addition to potentially affecting prices and quantities, the agreements can also alter other competitive variables such as innovation, the quality and variety of the products, the cost structure or the discounts applied. The quantification of this harm is usually more complex due to the qualitative or difficult-to-observe nature of these variables.

18. The quantifiable harm caused by such conduct can be divided, in a very simplified way, into two categories: on the one hand, harm caused by exploitative conduct (typically a price increase); and, on the other hand, harm caused by exclusionary conduct.

19. **Anticompetitive conduct of an exploitative nature** (e.g., cartels and exploitative abuse, which often lead to price increases) results in predominantly vertical market effects. This is because the harm is generally concentrated on the purchasers (direct and, possibly, indirect) and suppliers (through lower sales volumes resulting from the cartel, or through lower prices in the case of a purchasing cartel), while competitors may benefit from the cartel[14]. This harm is usually grouped around two main dimensions[15]: the price effect and the volume effect.

   i. On the one hand, the **price effect** is the result of purchasers (or suppliers, in the case of conduct relating to these) having to pay higher prices (charge lower prices) for each unit of the product concerned than would be the case in the absence of the infringement; it is often identified with the so-called **actual loss**.

   ii. On the other hand, the **volume effect** occurs when a purchaser (seller) of the product concerned[16] passes on part of the overcharge to their purchasers, resulting in lower sales and consequently lower profits compared to the non-infringement situation. This type of harm is often identified with the so-called **loss of profit**.

20. **Exclusionary conduct**[17], on the other hand, generates harm that primarily operates horizontally, affecting competitors at one or more levels of the

[14] In the event of a cartel leading to an overcharge, non-participating competitors may benefit because the cartel's supra-competitive price would allow them to set a higher price than would have been the case under free competition conditions. This is known as an umbrella effect and it harms buyers of non-cartelised products.

Although claims based on these "umbrella effects" are rare, EU legislation recognises the right of the injured parties to claim damages from the members of the cartel that caused the harm (see Section 33 of the Judgement of the CJEU of June 5, 2014, in case C-557/12 Kone AG et al.), unlike that which happens in other jurisdictions, such as the United States.

[15] As indicated, they can also give rise to other forms of harm (quality reduction, variety reduction, innovation reduction, etc.). Although these are more difficult to quantify or observe, this does not prevent them from being accredited and quantitatively approximated in specific cases. This Guide does not address such additional forms of harm.

[16] This purchaser cannot be a final consumer but must use the cartelised product for their commercial activity. For example, it may be resold or used as an input in their productive activity.

[17] Its specific features are discussed in Section 2.5.

value chain, generally leading to a loss of profit. It may also harm consumers and suppliers due to price changes[18].

21. Finally, to understand how harm is generated, it is crucial to understand the **interconnection between the actors involved** at different levels[19], the following agents can be distinguished:

  i. **Infringing parties**[20], who engage in any of the behaviours described above.

  ii. **Purchasers**, distinguishing between **direct purchasers**, who buy directly from infringers, and **indirect purchasers**, who buy a product affected by the infringement from direct or other indirect purchasers. Either group may include the final consumer, depending on where in the value chain the infringement has occurred.

  iii. **Suppliers** that supply the infringing parties (whose business may be negatively affected by the infringement due, for example, to lower sales volumes caused by overcharging), or who had been supplying to the foreclosed competitors.

  iv. **Competitors**, both companies affected by exclusionary conduct (this can range from exclusionary abuses to agreements between companies with foreclosure effects), and potential competitors who are prevented from entering the market because of barriers to entry imposed by the infringer.

  v. **Other agents:** for example, the producers of complementary goods may also be harmed if the infringement leads to a decline in sales of the affected product.

2.1.2. Passing-on of overcharge

22. Situations of overcharge pass-on arises when the harmed agent competitor, supplier or purchaser) passes on part or all of the harm suffered to their direct purchasers. Consequently, the agents who initially suffered the harm will, by passing it on, see the harm suffered reduced or even eliminated. In these cases, the quantification of the harm (and, therefore, the amount of compensation) must also be adjusted according to the degree of passing-

18 Although the prices paid by consumers may initially be lower, they will tend to be higher at a later stage if the behaviour leads to a reduction in the competitive intensity of the market. Competitive variables other than prices, such as quality, variety, or innovation, could also be affected.

19 The example could be replicated for a case where the infringement impacts the suppliers.

20 Mainly companies, although they can also be business associations, professional associations, etc.

on. On the other hand, indirect purchasers located at different points in the supply chain are entitled to claim damages[21].

23. The issue of passing-on overcharge has already been studied by the Supreme Court in relation to the **sugar cartel**, where some of the most controversial issues have been analysed (e.g., how to estimate this concept)[22]. The ruling establishes that the price increase by the purchaser directly affected by the conduct is necessary, but not sufficiently serious for the overcharge to be passed on. It is therefore necessary to prove that this price increase to its customers passes on the harm suffered due to the "upstream" price increase[23]. With the **amendment of the LDC resulting from the transposition of the Damages Directive**, and as already mentioned in Section 1.3, it has been clarified that the right to full compensation only covers the overcharge that the injured party has effectively had to bear, i.e., that they have not passed on or transferred to other members of the value chain, such as their own direct customers or the final consumer. However, the compensation for the actual loss suffered at the different levels of the value chain may not, in any case, be greater than the actual overcharge incurred at that level, since otherwise there would be overcompensation or unjust enrichment on the part of the plaintiff. The latter does not preclude that there is also a right to claim for loss of profit as a consequence of the total or partial passing-on of the overcharge[24].

24. Furthermore, in line with the above, the existence of downstream cost pass-on by the plaintiff can serve as a defence for the defendant, as Article 78.3 of the LDC recognises this possibility, shifting the burden of proof. This possibility may lead to a complete exoneration of the defendant or serve as a modulating element in terms of the possible compensation that they may have to pay since the LDC recognises the possibility of passing-on or transferring price increases to subsequent links in the supply chain. Thus, the harm suffered is equal to the overcharge actually paid minus the passed-on overcharge, plus the loss of profit suffered as a result of the volume reduction.

[21] As indicated in paragraph 14 and footnote 14 of the "Guidelines for national courts on how to estimate the share of overcharge which was passed on to the indirect purchaser" of the European Commission, the concept of passing-on harm could also be applied to direct and indirect suppliers, especially in relation to the volume effect.

[22] STS 5819/2013, dated November 7, 2013, ECLI:ES:TS:2013:5819. This judgement is based on the Decision of the Competition Defence Tribunal of April 15, 1999 (File 426/98, Sugar).

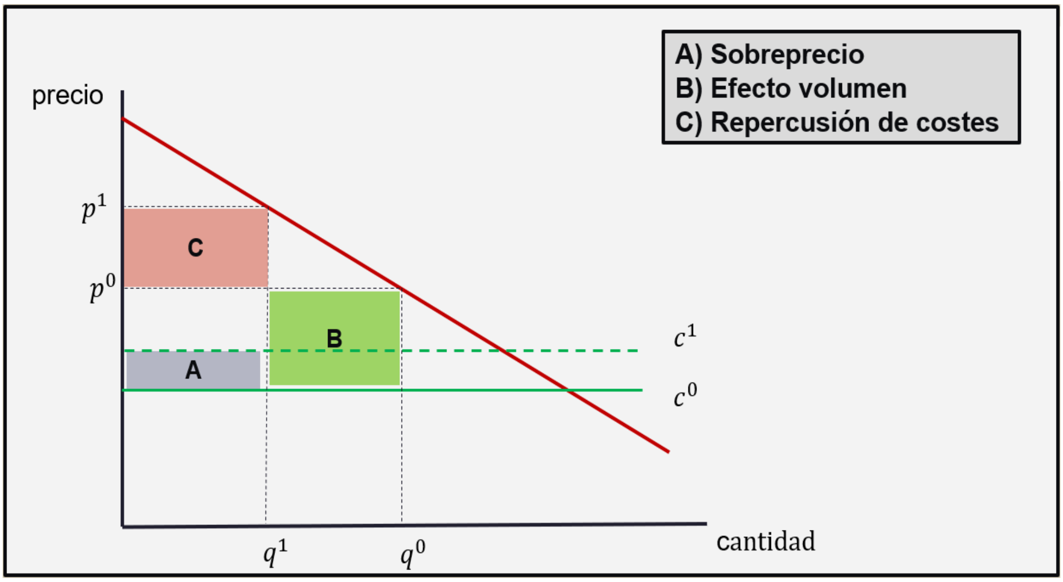
[23] In this context, "price increase" should be understood as the difference between the observed prices and those that would be expected in the absence of competition infringement (i.e., the counterfactual prices).

[24] These issues are included in Article 78.1 of the LDC.

25. **In economic terms**, the concept of overcharge pass-on and the different elements of harm are illustrated in Figure 1. The graph represents a hypothetical situation of a company that suffers anticompetitive conduct from its suppliers and partially affects its customers[25]. Firstly, the company, as a direct purchaser of a product affected by an infringement, suffers an overcharge that causes an increase in its marginal costs from c_0 to c_1. Secondly, as a consequence of its higher costs, the company increases its own sales prices to customers (from p_0 to p_1). Due to the higher sale price, the company experiences a reduced quantity sold (from q_0 to q_1).

26. Ultimately, the harm suffered by the direct purchaser of the products that are the subject of the infringement can be seen as the sum of area A (overcharge borne) and area B (volume effect) minus area C (overcharge passed on)[26].

Figure 1. Illustration of overcharge pass-on



Source: prepared in-house.

27. With regard to the downstream **links** in the supply chain, Article 79 of the LDC expressly entitles the indirect purchaser to bring an action for damages, although it imposes on them –as a general rule– the burden of

[25] For simplicity, it has been assumed that the company operates in a market with imperfect competition and that it has a linear cost function in which marginal costs are constant.

[26] If the direct buyer were a final consumer of the affected product, the total harm would be equivalent to the sum of the overcharge incurred and the irrecoverable loss of efficiency. This last component of the harm would be the loss of utility experienced by those consumers (actual or potential) who would have bought a greater quantity of the product if the price had been that of the non-infringement scenario.

proof that the overcharge has been passed on to them and that they have suffered harm.

28. It is, therefore, necessary to differentiate between the **two basic cases** in which the passing on of the overcharge can be brought before the courts:

- **Defence of the infringer** – defendant– against claims made against them by a direct or indirect purchaser – claimant – (Article 78.3 of the LDC). In this case, the burden of proving that the overcharge was passed on is on the defendant, who may reasonably require the production of evidence in the possession of the plaintiff or of third parties. This is known as the defensive or "shield" aspect of cost pass-on.

- **Basis of action of the indirect purchaser** – plaintiff– against the infringer – defendant– (Article 79 of the LDC). In this case, the burden of proving that the overcharge exists and was passed on is on the defendant, who may reasonably require the production of evidence in the possession of the defendant or of third parties. This is known as the offensive or “sword” aspect of cost pass-on.

29. In 2019, the European Commission published the "*Guidelines for national courts on how to estimate the share of overcharge which was passed on to the indirect purchaser*". Based on this document[27], **several factors** can be highlighted that, according to economic theory, may affect the existence and magnitude of the pass-on[28]:

- The **demand characteristics** faced by direct purchasers, particularly, the **sensitivity of demand to price changes (price elasticity)**. The more inelastic the demand, the easier it is to pass on overcharges, as the quantities demanded will react weakly to price changes (e.g., in the case of essential goods). In other words, the smaller the reduction in quantity demanded associated with a price increase (volume effect), the more likely it is that the overcharge will be passed on.

- **Intensity of competition in the market where direct purchasers operate and proportion of the market affected.** The impact of competition on the degree of pass-on depends to a large extent on whether the initial overcharge affects only the direct purchaser or

[27] In particular, section 3, “The economic theory of passing-on”. These issues are covered in greater depth in Annex 1.

[28] It should be noted that the order of appearance of the factors does not prejudge their relative importance with respect to the rest and they must be assessed as a whole (taking into account possible interrelationships) for each particular case.

whether the direct purchaser's competitors are also affected. When the overcharge only affects one direct purchaser, the more intense the competition faced by this purchaser, it is less likely to have an impact. Conversely, if the overcharge also affects the direct purchaser's other competitors, more competition between them tends to result in more pass-on compared to less competitive environments.

- The **cost structure** of the direct purchaser affected by the practice. Specifically, if the overcharge affects **fixed costs**, pass-on is less likely than if it affects **variable costs**, which are closely related to marginal costs[29], at least in the short term. Another related factor is that **the longer the duration of the anticompetitive conduct**, the more likely it is that there will be cost pass-on. This is because, depending on the time horizon, some of the fixed costs may become variable.

- The relevance of the input affected by the conduct. The greater the **relative importance of the input affected** by the anticompetitive conduct in the **price of the final product**, the more likely it is that the cost overcharge will be passed on.

- Costs associated with price changes. If direct purchasers face **costs associated with price adjustments**[30], it may not be worthwhile to pass on the overcharge unless the magnitude of the overcharge is high.

- The **countervailing power** of demand. If direct purchasers themselves have buyers with strong bargaining power, they may be constrained when it comes to passing on overcharges downstream. Other factors affecting negotiations between purchasers and sellers at different levels of the value chain also should be taken into account.

- The existence of **price regulations** affecting the product sold by direct purchasers may limit the degree of pass-on.

- The relationship between the affected product and other products sold by the direct purchasers. The direct purchasers may sell **other products whose demand is related to the product affected** by the overcharge. In particular, if they sell substitute products, the pass-on of the overcharge will increase the demand for such products, thus

[29] According to economic theory, marginal costs are relevant for companies when setting prices to maximise their profits.

[30] For example, the costs associated with printing new catalogues, renegotiating the price with customers or lost sales if prices are set just below an appropriate psychological level.

providing an incentive for pass-on[31], while the reverse is true for complementary products.

2.1.3. Access to data

30. Damages proceedings are characterised by the existence of **asymmetric information between the parties**[32]. It is reasonable to assume that the party who has been harmed by anticompetitive conduct may lack sufficient means of proof to demonstrate the harm suffered and its extent. For this reason, the law must provide them with the appropriate means to enable the effective exercise of their right to make a claim.

31. RDL 9/2017 acknowledges this difficulty and regulates access to sources of evidence by introducing Section 1 bis, in Chapter V of Title I of Book II of the LEC, entitled "*On access to sources of evidence in proceedings for damages claims for infringement of competition rules*".

32. Within this regulation, namely in Article 283 bis i), specific provisions are included on **access to sources of evidence found in the files of a competition authority**. This possibility granted in jurisdictional proceedings should not be confused with public access to the administrative file, which is regulated by Law 19/2013, of 9 December, on transparency, access to public information and good governance (hereinafter, Transparency Law). On numerous occasions, individuals who have suffered harm as a result of an infringement sanctioned by the CNMC and who intend to pursue their claims before the courts, resort to the Transparency Law to base their claims for access to the file before the CNMC.

33. However, recourse to the Transparency Law has well-defined limits that sometimes prevent claimants from satisfying their demands to obtain sufficient evidence through this channel to articulate their claims. The second section of the first additional provision of the Transparency Law establishes that "*those matters that have a specific legal system for access to information shall be governed by their specific regulations, and by this Law in a supplementary capacity*". To this end, Articles 42 and 43 of the Law on the Defence of Competition contain a specific system for access to information, distinct from the general right of access to public information, files and registers.

[31] For example, if a retailer is affected by an overcharge on its coffee purchases, a substitute product could be tea, while sugar would be a possible complementary product.

[32] Although there may be exceptions, the defendant usually has more data to quantify the alleged harm, while the plaintiff is the one who can best quantify the impact of the harm and the volume effect.

34. The Council for Transparency and Good Governance has issued several rulings concerning access to CNMC files on several occasions (among others, in the Resolution of 15 September 2015, and 25 August 2017), stating that all information or documentation obtained by the CNMC as a result of its inspection work enjoys the status of reserved information under express legal mandate and that the first additional provision of the Transparency Law recognises the prevailing application of its specific regulations to matters that have a specific information access system[33].

35. For this reason, in order to adequately prepare the claim for damages, the injured parties have at their disposal the regulation of access to sources of evidence contained in the LEC, specifically in Articles 283 bis a) to 283 bis k).

36. Three aspects of the **regulation of access to sources of evidence** should be highlighted.

37. Firstly, **access can take place at different** stages of **the judicial process**: it can be requested before the initiation, in the application or during the course of the proceedings. To determine its admissibility, judges must assess the plausibility of the claim, the relevance of the evidence to be used and its proportionality, taking into account the cost of disclosure. In other words, the claimant must present all relevant evidence to support the plausibility of initiating legal proceedings for the recovery of damages resulting from an infringement of competition law and specifically propose the evidence they intend to use. If the court deems that the request is duly reasoned, it may agree to disclose evidence that is available to the defendant or a third party, although it must limit access to those aspects that are proportionate, avoiding indiscriminate information searches.

38. Secondly, **judges may require the competition authorities to disclose the evidence in their files** in cases where no party or third party is able, to a reasonable extent, to produce it (Art. 283 bis i.10). For their part, the competition authorities may express their position on the proportionality of the request for disclosure that they have received. In any case, Article 283 bis i) sets boundaries for divulging evidence contained in the files of the competition authorities, which are categorised into three lists:

    - A *blacklist*, which contains the types of documents that may not be disclosed under any circumstances. This list includes a) statements issued in the framework of a leniency programme and b) settlement

[33] File R/0147/2015 of September 15, 2015, and File R/0255/2017 of August 25, 2017.

applications. To reinforce this limitation, the LEC declares that evidence falling under this category, which is solely available to the parties by virtue of their access to the file, cannot be used in damages claims.

- A *grey list* consists of a) information that was prepared specifically for the competition authority's proceedings, b) information that the competition authority has prepared and sent to the parties during the proceedings, and c) settlement submissions that have been withdrawn. The limitation on the disclosure of this category of evidence is temporary: the courts may only order its disclosure after the competition authority has closed its proceedings.

- A *whitelist*, which includes evidence in the file that does not fall into the categories identified in the black and grey lists.

39. Finally, the judges must establish the mechanisms necessary to **protect confidential information without prejudicing the effectiveness of the proceedings**. It will be up to the parties to argue, before the court, to what extent they consider the information requested to be confidential. This analysis, which will depend on the circumstances of each case, will involve weighing the protection of business confidentiality and other principles such as the right to be heard, the right to defence, and the effectiveness of the proceedings.

40. The **European Commission (2020)** in its Communication on the protection of confidential information by national courts in proceedings for the private enforcement of EU competition law provides some useful criteria that can be applied by national judges when assessing the confidentiality of information[34]. In particular, the Communication considers as confidential any information that meets the following cumulative conditions:

- It is known only to a limited number of people.

- The disclosure of such information may cause serious harm to the person who provided it or to third parties, and it includes information that has commercial, financial or strategic value. Time is also an important consideration so that recent information and future prospects are more likely to be considered confidential than older information that may have lost its commercial value over time.

[34]Although it is framed in a different context, the European Commission (2015) also published a guide on *Best Practices* for the use of data rooms in procedures set out in Articles 101 and 102 of the TFEU and in the EU Merger Regulation (own translation).

- The interests that the disclosure of confidential information may harm are objectively worthy of protection. For these purposes, the reputational interest in the face of a damages award cannot be considered an interest worthy of protection.

41. It is also worth referring to the **CNMC Guide (2020)** on the handling of confidential information and personal data in antitrust proceedings under the LDC, which is based on a three-fold analysis:

    - The **nature of the information**, i.e., whether it constitutes a trade secret because it contains information the knowledge of which could effectively cause significant harm.

    - Its **dissemination**, i.e., whether there has been prior knowledge of that information by the parties in dispute or by third parties.

    - Its **necessity** in the procedure, in the determination of the facts or in the analysis and assessment that is the subject of the procedure, as well as in the right to defence of an interested party.

42. Among the **mechanisms that judges can use to protect the confidentiality of the data**, without prejudice to their effective application to produce their quantification, the above-mentioned European Commission Communication, which regulates the following aspects in detail, is very useful[35]:

a) Censorship (or the unbundling of information)

43. In certain cases, the disclosing party may be required to redact copies of the documents by deleting confidential information. This mechanism may be advisable in cases where:

    - The volume of information is not too high.

[35] It should be noted that some judicial bodies have already drawn up their own rules for handling the business secrets and confidential information contained in the claims they process. Such is the case of the Protocol for the Protection of Business Secrets, of November 2019, prepared by the competition section of the Commercial Courts and Tribunals of Barcelona. The Protocol provides for various types of measures to protect confidential information, including locked custody in court premises, no direct transfer of copies, digital security measures and confidentiality circles, confidential and non-confidential versions of the parties' input and of the judgement, and restrictions on the publicity of oral hearings and access to recordings.

- The redacted versions are still meaningful and adequate in the exercise of the rights of the party requesting the disclosure.

- The holder of the information is a third party other than the party making the disclosure, and they should be required to indicate what part of the information they consider confidential to the requester.

b) Confidentiality circles

44. Through this instrument, the disclosing party grants access to certain categories of information, including confidential information, available only to designated individuals. This would encompass various mechanisms such as data rooms. Thus, by limiting the number of people who can access the information, the potential harm to the disclosing company is minimised. Confidentiality circles, further developed in Section III.C. of the European Commission Communication, may be effective in cases where:

    - There is a large volume of information, so censorship would be very costly.

    - The nature of the confidential information makes it very difficult to summarise without it losing its meaning and probative value.

45. If the creation of a confidentiality circle is deemed appropriate, to preserve its effectiveness and considering the proportionality of the proceedings to the intended purpose, it is important that basic principles are clearly established regarding:

    - **Accessible information**. The categories of information or specific items of evidence to be disclosed should be determined *a priori*, as precisely as possible.

    - **Composition**. The court must determine who will form the confidentiality circle, depending on the circumstances of each case and the nature of the information to be disclosed. In particular, it must determine whether, in addition to the parties and their legal advisers, other internal or external advisers may be present.

    - **Confidentiality commitments of the circle members**. Such commitments may include a duty not to display or disclose the confidential information obtained, or an obligation not to use it outside the process in question, a commitment to destroy or return any copies obtained, or a time limit on access to the information.

- **Logistical arrangements**. It can be decided whether the disclosure is to be made physically or digitally, when and where it is to take place, as well as how the documentation is to be made available.

c) Appointment of experts other than party-appointed experts

46. The aim is for them to have access to confidential information and provide a concise summary of it. This approach can be useful when handling commercially highly sensitive information of a quantitative or technical nature.

47. **Other mechanisms for protecting confidential information**, once it has already been collected and fed into the process, may include: i) the holding of hearings *in camera* or with restricted access; ii) the preparation of confidential and non-confidential versions of the judicial decision, without prejudice to the parties' right of defence; and iii) the total or partial limitation of access to the file at the judicial site.

48. It should be noted that all the mechanisms for protecting confidential information that have been mentioned, and which are described in detail in the European Commission's Communication, are compatible with the measures that, for the same purpose, are set out in Article 283 bis b) of the LEC, since both texts emanate from the Damages Directive.

2.2. The expert report

49. One of the most important elements of the damages claim procedure is the expert reports or rulings[36]. These reports provide the information necessary to prove, where appropriate, the existence of harm and its causal link with an infringement of competition law, as well as quantify the corresponding harm and facilitate the proper assessment of the quantification exercise by the competent judicial body.

50. Both the expert reports provided by the plaintiff and the defendant should **facilitate the understanding of the arguments presented and the replicability of their results** through a clear description of the data and methodology employed. Therefore, it is desirable that the drafting of the expert reports followed the **same principles** previously highlighted by the CNMC in the context of the economic reports submitted to the Competition Directorate (CNMC, 2018). These principles are:

[36] The object, purpose, evaluation and timing of submission of expert reports, among other issues, are regulated in Articles 335-352 of the LEC.

- **Completeness:** including all the information necessary **to be understood, reproduced** (e.g., by the other party to the proceedings if it so wishes) **and evaluated** by the court.

- **Transparency:** to encourage **understanding of the results** and **the replicability of calculations**, it is highly recommended that the expert report incorporates, in a transparent manner and in as simple a language as possible, the **data used and its processing**, the **assumptions applied** (explicit and detailed) and their **justification**.

- **Consistency:** the assumptions and results of all the analyses contained in the report should **be consistent, without any contradictions**. If there are inconsistencies, these should be explained, and it should be indicated which one is considered dominant.

51. The degree of detail of each expert report **will depend on each case** (data or time available, as well as material, human, financial resources, etc.). Some reports will incorporate a more descriptive and qualitative analysis and others will be more quantitative, ranging from simple statistical techniques to more advanced econometric analysis. Despite the differences between approaches, **they can all be equally valid**, as long as they accurately capture reality and present a justified and valid counterfactual scenario. Regardless of the approach used, the reports should have a common denominator: they should demonstrate the rigour of the harm quantification.

52. The following is a set of **best practices** regarding the content of expert reports to ensure they are as explanatory as possible. It should be emphasised that in no case should this be understood as an exhaustive list of what a report should contain. Indeed, it is up to the court to consider in each case whether the information contained therein is sufficient to adequately quantify the harm. The recommended best practices should be assessed from the perspective of the principle of proportionality and are desirable to the extent that they do not impose a disproportionately high standard on plaintiffs.

2.2.1. Characteristics of the affected sector and market

53. To fully understand the development and effects of the anticompetitive infringement, it is essential to have a **good knowledge of the sector and to reflect it adequately in the expert report**. This would include a discussion and analysis of any characteristics considered likely to affect the variable being quantified. Among others, one could consider:

- The **geographical scope** of the affected market or markets (local, regional, national or international) and the **applicable regulations**.

- The **determinants of supply and demand for price formation**. Among the determinants of supply, we can highlight the **cost structure** of the companies (main factors of production used and their cost, economies of scale, proportion of fixed and variable costs[37]). Among the determinants of demand, we can point out consumer preferences and the price sensitivity of demand for the affected goods or related goods (substitutes and complements)[38] and for the changes in purchasing power[39].

- The **maturity level of the affected market**, differentiating between those markets that are in a growing and rapidly evolving phase and those that are established and mature or, where appropriate, in decline.

- The level of **competition existing** in different segments of the affected sector, where data on the number of competitors and their market share, together with the frequency of new entry and exit of companies, the existence of high entry costs or the degree of product differentiation may be particularly relevant.

- The determination of whether the affected market deals with **intermediate or final products**, in order to narrow down the potential **pass-on of cost overcharges** between the different stages in the value chain. The form of marketing through the different stages may also be a relevant factor.

- The **interaction dynamics between sellers and buyers** in the relevant market. This may include aspects such as the pricing process (e.g., whether list prices are applied or individually negotiated or whether discounts are available), the duration of contracts or the costs of switching suppliers or customers. A distinction should be made between traditional markets or tenders (where interactions tend to occur less frequently, so that anticompetitive effects may be prolonged over time depending on the duration of tendered contracts).

---

37 In the short term, price variations tend to be influenced to a greater extent by variable costs than by fixed costs.

38 The price elasticity of demand shows how the demand for a product varies in response to changes in the price of that same product (own-price elasticity) or of a competitor's product (cross-price elasticity).

39 The income elasticity of demand shows how the quantity demanded of a good varies with changes in consumer income (without changing prices).

54. All these characteristics directly influence the selection of the applicable analytical methodology, and it is necessary to understand their key dynamics to correctly quantify the harm.

2.2.2. Theory of harm and description of the counterfactual

55. It is important for the expert report to **provide an objective and coherent description of how the anticompetitive conduct may have led to the specific harm** being quantified (the theory of harm). If the defendant's expert report concludes that there is **no harm**, an economic rationale for the facts found and of the non-existence of such harm should be presented[40]. Although the plaintiff usually refers to how the harm suffered has occurred, it should also be a fundamental part of any expert report, as it forms the basis for constructing the counterfactual scenario and for properly quantifying the harm.

56. The counterfactual scenario is the **hypothetical situation that would have occurred** in the absence of the anticompetitive infringement. Therefore, to quantify the harm, the actual situation of the injured party must be compared to that hypothetical situation. For this, it is essential to properly understand the nature of the anticompetitive practice, as well as the sector and market affected by it, as indicated above.

57. The design of the counterfactual scenario is a complex exercise, since it is entails establishing how the agents would have interacted and what the market conditions would have been like in a scenario that is not observable because it has not occurred in reality (i.e., it is hypothetical). Its design, therefore, is not free of uncertainty, as it involves making several assumptions to reflect a situation as close as possible to that which would have arisen in the absence of the infringement. In this way, it aims to isolate the effect of the anticompetitive conduct.

58. A crucial issue when constructing the counterfactual is the **temporal delimitation of the damages claimed.** This exercise must be carried out in a reasoned and transparent manner, in the same way as when trying to determine the magnitude of the harm in each period, as this will substantially affect the quantification.

59. In the case of follow-on complaints, the duration of the infringement set out in the decision made by the competition authority could be a good starting point, considering the possible extension of this initial duration that could be

---

40 This explanation is always required in cases of cartels, given that Article 76.3 of the LDC stipulates that "infringements classified as cartels will be presumed to cause harm unless proven otherwise."

accredited in subsequent surveillance decisions[41]. However, there are several reasons why, from the quantification of harm perspective, the period during which the effects of the infringement occur may not necessarily coincide with the duration of the infringement stated in the administrative decision. In these cases, a detailed justification of the divergences is necessary. The infringement may have commenced earlier than stated in the decision or, once it had begun, it may have taken some time to take effect[42]. It is also possible that the effects of an infringement may continue despite the cessation of the practice, sometimes referred to as a "lag effect"[43].

60. In cases where it is difficult to clearly delineate the temporal effects of the infringement, it is even more important to justify the choice of the infringement period transparently, assessing possible scenarios and, foreseeably, analysing the sensitivity of the results to the different alternatives. If this uncertainty occurs, one possibility is to omit the periods close to the infringement when constructing the counterfactual scenario, although this approach has several drawbacks[44]. Graphical and descriptive analysis techniques can be useful for detecting the beginning or end of the effects. Additionally, where the data so permits, a temporal delimitation can be carried out using econometric techniques. An example would be to use a structural change test on the variable analysed, which allows the identification of when there was a change in the behaviour of the variables and the relationships between them without the need to make prior assumptions about their causes.

61. The construction of the counterfactual scenario is a key point in the analysis, given that its erroneous specification could significantly affect the harm quantified, potentially over- or underestimating the real effects. The **expert report should therefore provide a detailed and transparent explanation of the criteria used for its construction** and the factors that could lead to

41 See, in this regard, Article 41 of the LDC and Articles 42 and 71 of Royal Decree 261/2008 of February 22, which approves the Regulations for the Defence of Competition.

42 For example, if there are deviations from the participants in the infringement with respect to what was agreed (in the case of restrictive agreements) or if there are contracts or regulations causing delays in the effects to materialise.

43 This may be due to the existence of long-term contracts or other factors that introduce rigidity in the variables of interest, prolonging the effects of anticompetitive conduct. In cartel cases, the effects may also be prolonged by tacit coordination between offending companies.

44 On the one hand, the more we move away from the infringement period (backwards and/or forwards in time), the more likely it is that other factors will appear that influence the variable of interest and that, if not adequately controlled, make it difficult to assess the effects of the infringement (RBB Economics and Cuatrecasas, 2017, para. 393). In addition, it must be considered that, even if it is not clear if the periods close to the infringement were affected by it, they can be incorporated into the analysis to obtain a lower limit for the harm (para. 46 of the EC's Practical Guide).

an over- or underestimation of the real harm, as well as an indication of why other possibilities for the counterfactual were rejected.

62. Defining a scenario that reflects what would foreseeably have happened if the infringement had not occurred, is not easy. However, it is the expert's task to choose the appropriate assumptions and hypotheses, which can be accepted as sufficiently reasonable and probable (and not as unjustified choices that decisively influence the results obtained), and which allow the construction of a scenario that is sufficiently similar and approximate to what would have occurred without the infringement.

2.2.3. Selecting the relevant variables and data used

63. Two other very important issues to take into consideration in the expert report are the **proper selection of the relevant variables** and the **construction of the database** used to carry out the analysis. These issues, in turn, will be influenced by the approach adopted[45].

64. Once the theory of harm has been explained, one would expect an expert report to describe which variables should be used to quantify the effect of the conduct. These variables should be determined on a case-by-case basis, as quantification depends on the nature of the infringement, the sector, market, and product affected, the applicable sectoral regulation, the legal framework in place and the characteristics of the claim. To analyse follow-on claims, a good starting point for analysing the characteristics of the market and selecting the variables affected by the infringement may be to refer to the competition authority's decision.

65. Among the **most commonly used variables for quantifying harm** are prices, sales volume, business margins and profits. When calculating the harm, it is necessary to consider the effect of factors unrelated to the anticompetitive conduct on variables of interest[46]. For example, if the price level is selected as the variable for estimating the harm, it should be noted that prices in the actual (with infringement) and counterfactual (in the absence of infringement) scenarios may differ, **not only because of the infringement but also because of other variables that should be taken into account when quantifying the harm**. These variables could be, for instance, costs (e.g., price of raw materials, labour costs), utilisation of production capacity, factors that approximate the degree of competition (e.g., number or market share of competitors, imports), substitute and complementary products (e.g., prices and quantities sold), factors

[45] See, in this regard, the following subsection.

[46] When resorting to econometric techniques, these external factors are called control variables.

influencing demand (e.g., income, population, production of goods containing the affected product, consumer preferences), variables describing relevant macroeconomic developments (e.g., GDP developments) and possible disruptive elements, such as technological shocks, quality improvements or regulatory changes[47].

66. Once the variables considered most appropriate for quantification have been described, the possibility of obtaining these, their characteristics[48] and the sources to be used for this purpose should be addressed.

67. Publicly accessible data sources, such as the websites of organizations like INE, Banco de España, Eurostat, OECD, and World Bank, among others, which publish cleaned data, in the form of time series related to production at sectoral level and to the macroeconomic evolution of different geographical areas (national, regional or local data). These data sources can be very useful for quantification: (i) they are easy to obtain, (ii) they include information from agents unrelated to the infringement, allowing comparisons between different companies, and (iii) any biased use of them would be easier to identify, since the data is accessible and traceable to the parties. However, public sources are often not sufficient to construct damages claims, which by nature must focus on the study of the interaction of agents at the individual level.

68. For this reason, it is often necessary to resort to **private data sources**, as these allow for greater disaggregation of the information and are, therefore, usually better adapted to the specific case. However, obtaining them can be time-consuming and costly, and present a problem of asymmetric information: in most cases, it is the defendants and third parties who have the necessary data for an accurate and robust analysis[49]. In any case, the use of private sources should not be a problem as long as it is possible to obtain detailed knowledge on the database, the original source, the treatment applied and any limitations.

69. It must be considered that the **availability of data** can greatly condition the selection of variables and, ultimately, the quantification of the harm. Therefore, although access to **good quality and highly reliable data**

[47] It should be noted that economic theory does not establish a list of variables that determine the supply and demand of a product in all cases, instead, this depends on the specific case. Hence the importance of adequate knowledge of the sector affected by the conduct.

[48] This would include issues such as the type of information included, the units of measurement and the periodicity.

[49] On other occasions, for example, to quantify the pass-on or volume effect, the direct customers usually have more information.

should be a fundamental objective before undertaking a quantitative analysis, in practice it is not uncommon that some of the variables of interest cannot be obtained[50] or the databases are incomplete and require specific treatment (e.g., outliers, imputation of missing values, aggregation)[51]. In such cases, the decisions taken should be motivated and transparently explained.

70. To quantify harm, parties in the process typically access a significant amount of data. However, it is desirable for both sides to have access to the data in a format that can be processed by computer applications, as well as the codes, commands or programming procedures used in the analysis, **in order to reproduce or challenge the conclusions**.

71. In addition, the information gathering has to facilitate **the construction of a case-specific claims database** containing a **sufficient number of observations** to ensure robust results[52]. It is also advisable to incorporate **sufficiently representative and continuous time series**, avoiding the use of partial periods without due motivation.

72. It is highly recommended that the expert report includes a **descriptive statistics** section, providing information on, among other aspects, the following[53] (CNMC, 2018):

- **the data collection process:** the sources used, the method for obtaining the sample, the use of some data versus information, and so on.

- **the sample characteristics:** the number of observations, the units of measurement, the problems encountered (data not available, the existence of extreme values, etc.).

- **the treatment applied to the original database to arrive at the one used in the quantification**: interpolations, extrapolations, imputation of missing values, detection and treatment of extreme values of variables,

[50] In those cases, the variable of interest could be omitted or, alternatively, other variables (called "*proxies*") could be employed to approximate their behaviour. For example, in the absence of the cost function of the defendant company, price indices of the most relevant inputs in the production of the good affected by the infringement could be used.

[51] See, in this sense, Section A2.1.3 and Subsection A2.2.6.1 of Annex 2.

[52] The number of observations sufficient to carry out statistical inference depends on numerous parameters chosen by the analyst, among which it is worth mentioning the desired minimum detectable effect or the level of statistical significance. It is not common to find explicit analyses on the size of the sample in the economic literature on the estimation of harm for specific cases, this being more commonly mentioned in theoretical documents.

[53] See Section 1 of Annex 2.

deletion of records, aggregation of data to the same timeframe, sectoral processing, and so forth. In addition to detailing all the processing transparently, it is necessary to justify this and explain to what extent the representativeness of the sample has not been compromised.

73. All these issues are necessary for the replicability of the exercise which, in turn, results in greater transparency. Therefore, it is recommended they play an important role in the expert report.

74. In summary, it is recommended that the expert report includes a **descriptive section** on the **selected variables** (definition and description, including the units of measurement, justification of their choice or omission, description of the problems encountered, analysis of possible correlations between variables, etc.) and of the **data used and its processing**.

2.2.4. Techniques used and presentation of results

75. Along with the analysis of the sector and markets affected, the counterfactual scenario, the variables that most accurately and comprehensively allow the harm to be quantified, and the availability of data, the expert report needs to **address the methodology used** in the quantification[54].

76. A detailed explanation of the application of methods (dealt with in the next section of the Guide) is key to evaluate the results obtained. Quantification methods, which can be used to quantify both actual loss and loss of profit, often rely on the use of statistical or econometric techniques.

- **Simple statistical techniques** allow the data set used to be grouped into a series of descriptive values, such as the mean, mode, median, variance, standard deviation and coefficient of variation, among other things, which make it possible to understand the structure of the data and identify certain patterns. These descriptive values, usually accompanied by graphs, allow an approximation of the reality to be quantified. Descriptive statistics are often used when data availability is limited or as a starting point for more complex analyses, such as econometric analyses. This is because descriptive statistics usually do not allow the effect of the infringement to be isolated and therefore do not allow specific quantification of the harm resulting from the infringement.

54 The chosen method will depend on all the above and, in turn, will determine the selection of variables and data used.

- **Econometric techniques** (mainly regression analysis) typically involve more complex tools and require greater data availability. They are often used to quantify the underlying economic relationships and look for links between the variables used to measure the harm. To build a model or an econometric regression, an observable variable capable of capturing the effects of the infringement, i.e., the **explained, dependent or endogenous variable** (often prices, sales volume, margins or profit) will be determined. On the other hand, a set of **independent or exogenous explanatory variables** that are considered to influence the determination of the dependent variable will be established. Once the variables and the database have been selected, and the model or econometric regression has been designed, the estimation method will be chosen, the result of which will provide information on the differential evolution of the explained variable in the counterfactual scenario (where there is no infringement) and in the real or observed scenario (with infringement).

  When assessing the degree of robustness and reliability of the estimated econometric model, different issues must be taken into consideration, such as the importance of the **error term** (which includes all the information that is not explained by the independent variables used), the variation of the dependent variable that is explained by the model (coefficient of determination or R2), the **level of** statistical significance or whether the model presents problems of **endogeneity** or **heteroscedasticity**, among other factors[55].

  It is recommended that the **expert report details all these issues**, starting with the method used and ending with a justification of each of the assumptions and hypotheses adopted (variables chosen, estimation methods, error term, tests and checks carried out, statistical significance and confidence intervals associated with the estimated coefficients of the relevant variables, etc.).

77. The following section details the most common methods for quantifying harm. Naturally, the expert report should include a detailed and precise description of the method or methods selected, as well as their justification and any limitations encountered. The application of these methods will lead to the **final result**, i.e., the quantification of the harm. The expert report must include a detailed account of how this result has been reached and the **degree of robustness of the model used**. To this end, it is advisable to check the sensitivity of the results to changes in the most controversial or

[55] For more information on the methodological care of econometric models, see Section 2 of Annex 2.

subjective aspects using a **sensitivity analysis** and, if necessary, to **justify any differences**. A robust model implies that minor changes in the method used do not lead to significant changes in the conclusions. In addition, the presentation of a sensitivity analysis will allow a range to be established within which the estimated harm is plausible, within the same model used.

78. If the defendant's expert report concludes that there is no harm, an economic explanation of the facts found and of the non-existence of such harm should be presented[56]. Whenever data and resources permit, it would be advisable to present **several approximations of the selected method**, in the interest of the increased validity and reliability of the results. In cases where the results of the different approaches show significant divergence and the different assumptions applied in each method hinder comparison, the reasons for the differences obtained should be indicated. Likewise, one should also question whether the results obtained constitute a minimum or maximum value of the harm caused by the infringement.

79. In short, the more **complete, precise, detailed, transparent and consistent** the expert report or ruling is, the more comprehensible and assessable the final result of the analysis will be for the competent judicial body, and the easier it will be to identify the differences and contradictions between the reports of the different parties in the proceedings, allowing the judicial body the chance to formulate specific questions to the experts in order to determine the harm caused.

80. Finally, it should be stressed that quantification of harm based exclusively on harm **estimates from previous judgments should be limited to those cases** where it can be concluded that there is a sufficient degree of similarity. On the other hand, estimates based on the automatic application of an **average percentage** of previous infringements, or the **economic literature may lead to significant errors**, without prejudice to their being considered as **references**. Each claim, even if it involves the same conduct as another, may have particularities that require the quantification method to be adapted to the circumstances of the claim under consideration.

2.3. Methods for quantifying harm

81. Having analysed the key issues to be considered before selecting the quantification method, this section attempts to describe, without claiming to be exhaustive, the most common methods for quantifying harm. In any case, a harm quantification method other than those presented in this Guide

[56] This explanation is always required in cases of cartels given that Article 76.3 of the LDC stipulates that "infringements classified as cartels will be presumed to cause harm, unless proven otherwise."

should **not be rejected out of hand**, but it will be necessary to consider whether its development and presentation are in line with best practice and whether it produces a sufficiently accurate result, taking into account the particularities of each case and the constraints imposed by the information, resources and time available.

82. The different methods discussed in this section can be **complementary** as they reveal different perspectives (comparing periods, markets, using costs, simulations, etc.) and different levels of sophistication (simple statistical or econometric techniques, reduced-form or structural models, etc.). Sometimes, more than one method may be applied to the same case or, as is advisable, sensitivity analyses may be conducted, either in the same expert report or in the respective reports of the parties. If the **results are similar**, they may reinforce the conclusions or contribute to establish an estimate of the minimum level of harm caused by the infringement. On the other hand, if the **results are contradictory or differ substantially**, some authors offer two solutions (Oxera, 2009 and Seixas and Lucinda, 2019): i) selecting the approach that is considered preferable, primarily considering the chosen methodology and its application, available data or the robustness of the results (*best model approach*); or ii) combining the different quantifications, preferably omitting models with significant weaknesses (*pooling approach*) [57].

83. It should be borne in mind that it may be difficult to find appropriate criteria for comparing two methods with different approaches (e.g., a comparative method and a financial method) since each will have its own advantages and disadvantages intrinsically. The most important thing is, therefore, to analyse whether the chosen methods have been adequately appraised and to provide reasons for excluding the discarded methods[58].

2.3.1. Comparative methods

84. Among the most commonly used methods are **comparative methods,** which are based on (i) comparisons in the same market at a point in time before and/or after the infringement; or (ii) in a different but similar geographical market; or (iii) in a different but similar product market. A comparison of different time periods (diachronic comparison) may also be combined with a comparison of different geographical or product markets (iv). Therefore, before applying this approach, and as indicated in

[57] In general, it would not be appropriate to automatically calculate the arithmetic average of the proposed compensation or directly invalidate both results, but to analyse the causes that justify the possible divergences (para. 125 of the EC's Practical Guide).

[58] In this sense, the checklists in Section 2.6 of this Guide can be useful.

Subsection 2.2.2 of this Guide, it is necessary to justify the comparability of the markets or periods taken as a reference.

85. In its simplest version, **statistical techniques** are used to make the comparison. These have the advantage of contributing to the illustration of arguments about the theory of harm without requiring as much information as econometric techniques. However, the major drawback is that they do not allow for other factors that may have influenced the variable under analysis, i.e., they do not allow to isolate the specific effects of the infringement.

86. Another, more sophisticated possibility for isolating the effect of the infringement from other potential factors affecting the variable of interest is to use **econometric techniques**. These methods are usually based on **reduced-form models**, which assume equilibrium between supply and demand and attempt to condense the effect of the infringement on the variable of interest (the dependent or explained variable) into a single equation. The underlying assumption of reduced-form models is that, during the period under investigation, the fundamental economic relationships based on supply and demand are stable, and that the infringement did not lead to structural changes between the factual and counterfactual scenarios that cannot be accounted for by control variables.

87. In turn, reduced-form models are often applied to quantify harm in two ways. The **predictive approach** uses data from the situation without infringement to generate predictions of the variable of interest in the situation with the infringement. The ***dummy* variable approach** uses data from both situations and includes a dummy variable (whose value is generally 1 if the data corresponds to the infringing scenario and 0 otherwise) to reflect the differences between the two scenarios.

88. Reduced-form models are a simplification of **structural models**[59]. These models differ in that the latter attempts to explain the variable of interest by simultaneously considering the underlying economic relationships between the various supply and demand factors, such as the elasticity of demand or the cost structure of companies. Structural models have a solid theoretical foundation and can take into account changes in market structure caused by the infringement as well as other factors not captured by reduced-form models. However, their main drawback is that they generally require more data, and their construction and estimations are more complex.

---

59 Generally, these structural models are used as part of simulation methods (see Section 2.3.3 for more information).

89. The most widespread comparison criteria for contrasting actual results with a counterfactual are i) **comparison of different time periods or diachronic comparisons**; ii) **comparison of different markets or products (**also called **yardstick or synchronous comparison)**, including geographical or product comparisons; and (iii) a **combination of both**, either a market and period comparison, or the *difference-in-differences* method.

a) Comparison of different time periods (Diachronic comparison)

90. This method involves comparing the evolution of the variable of interest to quantify the harm during the infringement period with the evolution of the same variable in a period (i) before, (ii) after, or (iii) before and after the anticompetitive conduct.

91. This method has several **advantages**: (i) it requires data only on the product affected by the infringement in a single market; (ii) by using the same market and product, market characteristics (which are not always observable) are more likely to be comparable than when using different geographical or product markets.

92. However, the diachronic method also has certain **drawbacks**. Notably, the method indirectly assumes that the market structure has not changed during the periods under consideration. Therefore, it should only be used when this condition is met or when control variables that account for changes or relevant differences between the analysed period and the counterfactual[60] are included.

93. When comparing periods far apart in time, a mere comparison of prices (or other variables expressed in monetary units) could be biased due to inflation effects[61]:

- If quantification is based on simple statistical techniques, a possible solution would be to adjust the results using what is also known as "deflation". There are several price indices[62] available, and it is advisable

[60] Among the issues to consider could be fluctuations in demand, market seasonality, technological progress, or the existence of relevant changes (*shocks*) in the markets (e.g., a sudden increase in the price of a key input).

[61] The bias may also be due to other factors such as product characteristics, market technological evolution, new product development, or changes in market economic conditions.

[62] For example, the Consumer Price Index (CPI) is a statistical measure of the evolution of the prices of goods and services consumed by the population residing in family households in Spain; or the Industrial Price Index (IPRI), which measures the monthly evolution of the prices of industrial products manufactured and sold in the internal market, in the first stage of their commercialisation, that is, the sales prices at the factory, excluding transport and marketing expenses and VAT.

to use, as far as possible, those that are most closely related to the analysed product.

- On the other hand, if econometric techniques are employed, the usual practice is not to deflate the variables used, but to incorporate variables that may have influenced the evolution of prices (e.g., supply factors, such as costs, or demand factors) to try to isolate their impact.

In any case, the method used should be consistent and not mix variables expressed in nominal terms (i.e., without adjusting for inflation) with others in real terms (deflated).

b) Comparison of different markets and products (Yardstick or Synchronous comparison)

94. Market comparison can be of **two types**: geographical or product comparison. In both cases, the focus of the analysis assumes that the differences between the market affected by the infringement and the markets considered comparable are mainly due to the effects of the anticompetitive conduct.

95. A **geographical comparison** involves comparing the variable of interest during the infringement period with observations of that variable in the **same period** and **for the same product**, **but in another similar geographical market** that has not been affected by the anticompetitive conduct[63].

96. A **product comparison** consists of comparing the variable of interest during the infringement period with observations of that variable in the **same period and geographical market**, **for different products, but with similar characteristics**, which have not been affected by the anticompetitive conduct.

97. To correctly select the comparative geographical or product market, a comparative **analysis of its main market characteristics must be conducted beforehand**. Depending on the case, factors such as size and proximity (geographical and economic, among other aspects), characteristics of demand (income, population, among others) and supply (cost structure, nature and substitutability of products, among others), concentration and competition levels (number, characteristics and behaviour of competitors, among others), barriers to entry, regulation, maturity of the market, price-setting mechanism, or any other phenomenon

[63] The EC's Practical Guide recommends referring to the Commission Communication on the definition of the relevant market for EC competition law (OJ C 372, of 9.12.1997, p. 5) for further elaboration of the relevant market concepts (geographical and product).

relevant to the functioning of the markets could be taken into account. The more similar the market under consideration and the market affected by the infringement are, the more appropriate the comparison between the two.

98. This approach has the advantage of allowing the quantification when information is only available for the infringement period. However, the main disadvantage is that it can sometimes be challenging to find sufficiently similar product or geographical markets, as well as the necessary information required to control for factors that determine the variable of interest and that may differ between the two markets.

c) Difference-in-differences

99. The difference-in-differences method **combines time-based and market-based comparison**: it examines the evolution of the variable of interest (e.g., the price of the product) in the market concerned over a given period, encompassing the sub-period of the infringement together with a preceding and/or subsequent sub-period, and compares it with the evolution of the same variable over the same period in a comparable (geographical or product) market not affected by the infringement (as explained above).

100. This method helps to **isolate the effects of the infringement** from the influence of (i) factors that differ between the various comparison markets and have remained stable over time, and (ii) factors that have changed over time but affect both the affected market and the reference markets similarly. One advantage is that these factors can be taken into account without the need to explicitly include them in the analysis.

101. However, this methodology shares the **drawbacks** of the other comparative methods (difficulty in defining the periods and selecting similar markets) and may even require more information and data to be able to perform the double comparison[64]. Furthermore, this method requires the **assumption of parallel trends**, whereby the variable of interest would have evolved in the same way in the affected market and the comparison market in the absence of the infringement. To try to justify the fulfilment of this assumption, in addition to the relevant qualitative explanations, there are several options that focus on comparing the behaviour of the two markets over time (e.g.,

[64] While it is true that the difference-in-differences method requires the collection of the same explanatory variables in different markets (the infringement market and the relevant market not affected by the infringement) over long periods, as long as the markets experience parallel evolution in terms of the variable of interest, it could be valid to assume that variables not included in the models do not distort the results since they might evolve similarly in both markets.

graphical representation of the time trends of the main variables[65] and the use of simple statistical techniques or more sophisticated econometric methods[66]).

2.3.2. Cost-based methods and financial analysis

102. The objective of cost-based methods and financial performance-based analyses is to estimate a **reasonable and likely value** of the variable of interest (prices, profits, etc.) that would have resulted in the absence of anticompetitive infringement and to compare that value with the value actually observed for the same variable through costs or profitability.

a) The cost-based method

103. The **cost-based method**[67] is based on obtaining a **cost per unit** of production and adding to it a **"reasonable" business margin**[68], to arrive at a **"reasonable" price** that would have resulted in the absence of anticompetitive infringement. In this way, the harm would be the difference between the reasonable price per unit and the price per unit actually produced.

104. To apply this method, it is necessary to start by calculating the unit or average cost by dividing the relevant actual cost of production by the number of units produced of the product concerned. The necessary **cost data** for the analysis is likely to be contained in the company's accounting information, which may be either publicly available or internal. Adjustments to the accounting data may be necessary to arrive at the relevant unit cost for quantification. The adjustments and transformations made should be adequately explained and be in line with standard practice and the sector concerned. It is also possible to use non-accounting approximations to costs, in which case the method used should be explained in detail and the source should be properly referenced.

105. When applying the cost method, some particularities must be taken into account:

---

[65] It should be noted that graphical representation requires caution when illustrating trends in the variables and it is desirable to present a numerical verification.

[66] Annexes 2 and 4 set out some of these techniques.

[67] This is also known as the *bottom-up* method of costing or the *cost-plus* method.

[68] Three types of margins are generally used: gross profit margin, net profit margin and operating profit margin. The choice of one or the other depends on factors such as the type of business or the variables considered for the analysis (Oxera, 2009).

- **Publicly available accounting information** often does not have the level of disaggregation necessary for direct quantification. It may, therefore, be necessary to complement it with indirect sources[69].

- At times, **economic concepts do not coincide with accounting concepts**[70]. In case of discrepancies, the selected solution must be transparently explained.

- It is necessary to **justify the type of cost used** (e.g., variable, incremental, total), according to the characteristics of the specific case.

- Applying this method may be complicated in the case of **multi-product companies**, where only one of their products has been affected by the infringement. In this context, it may be complex to allocate common costs across the company[71] and the allocation method used should be explained in detail[72].

- Some types of conduct, such as cartels, may lead to a **reduction in the productive efficiency** of the participating companies due to reduced competitive pressure in the market and production constraints that lead to the underutilisation of economies of scale. Consequently, unit production costs may be higher than in a competitive environment. If there are indications that this is the case[73], it could be taken into account when determining the relevant costs, for instance, by using aggregate data from the sector or from similar companies or products not affected by the infringement. If econometric techniques are used, endogeneity can be addressed through instrumental variables[74]. Another option is

---

69 For example, publications from industry associations, specialised journals, or the international price of commodity price quotations can be used as approximations for input costs.

70 For example, the accounting profit is the difference between income and expenses in the income statement (they can be used to remunerate the owners or to increase reserves), while economic profit occurs when the company obtains income greater than necessary to compensate the opportunity cost of all the factors used (sometimes approximated as the change in the item of own funds from one year to another).

71 For example, certain raw materials, fixed assets, R&D, technology, or business services (IT, financial, legal, administrative, cleaning, logistics, etc.) may be common to various branches of activity of the company.

72 For a discussion of the different types of costs and methods for their allocation, see Oxera (2003), Section 6. The Decision of the European Commission in case AT.40394 – Aspen, paragraphs 108 to 115, may also be of interest.

73 To assess the probability that costs are affected by the variable of interest (for example, price), it is important to provide a detailed description of how and when the cost variable is determined, along with mechanisms to verify it if necessary.

74 See Annexes 2 (Subsection 2.5.2) and 4 (Subsection 4.1.3) for more detailed information.

not to make adjustments and consider the estimation as a minimum level of harm[75].

106. Several approaches can be taken to try to **approximate the reasonable business margin**. For instance, the possibilities include:

- A simple approach to "reasonableness" could be to compare the observed margin with certain statistics such as the **mean, median or mode**[76] **of the industry's business margin**, excluding infringing companies[77]. However, there may be reasons justifying differences in margins within a given industry, such as market share, consumer preferences, quality improvements, technological progress, and so on. Consequently, when calculating the fair value, the choice of companies used as a benchmark should be justified.

- An alternative benchmark could be the margin of the infringing company in a **period before or after the infringement**, or data on the operation of the same product in **different geographical areas or on other similar products** in the same geographical area, provided that the relevant scenarios have not been affected by the infringement. As with comparative methods, **the similarity** between the affected market and the reference markets (counterfactual markets), especially in terms of characteristics that most likely affect business margins (e.g., concentration and intensity of competition in the sector, cost structure, barriers to entry and exit, production capacity, economic cycle) should be explained.

- Thirdly, it is possible to infer a reasonable margin for the counterfactual by taking into account the structural **characteristics of the market** and trying to construct a hypothetical scenario in the absence of infringement using industrial organisation models[78]. This calculation requires flexibility depending on the sector concerned.

107. The **advantages** of using margins lie in eliminating the need to use other variables to control for changes in costs (e.g., inflation) given that the method incorporates information on the cost reported by economic agents

[75] In these cases, it is necessary to assess whether it is preferable to use the original variable, despite the potential difficulties that this may entail when quantifying the harm, or to use an alternative variable unaffected by the infringement but which may result in a less exact approximation of the relevant costs.

[76] See Annex 2.

[77] It may be useful to take into account the dispersion of industry margins when choosing the most appropriate indicator and analysing whether it is a good reference point.

[78] See Section 2.3.3 for more information.

and assumes that the effect of the infringement is reflected in the margins. Moreover, if the information comprising the margins (prices, costs, revenues) is of good quality and considered comparable, the use of margins may be a good estimator of overcharge, assuming that the infringement has not affected the costs.

108. Finally, in terms of possible **disadvantages** related to using margins to quantify harm, it may be challenging to find sufficiently detailed and reliable cost information. In addition, the infringement may also have affected the costs, which could lead to an underestimation of the potential overcharge[79].

b) Financial analysis methods

109. **Financial analysis methods** aim to approximate what the financial situation (usually profitability) of the defendant company or the plaintiff would have been in the absence of the infringement, as a benchmark for quantifying the harm suffered. They are especially useful for **loss of profit** claims.

110. The profitability of companies can be calculated in both monetary and percentage terms. Three techniques are commonly used to calculate company profitability (see Box 1 for clarification of the different financial and accounting concepts):

1. **Net present value** (NPV). This is one of the most widely used methods of assessing the value of companies and projects. It consists of calculating the aggregate value of the future cash flows of the company foreclosed by the infringement, discounted to the time when the infringement starts at a given interest rate. The NPV could provide an approximate value of the harm caused to the foreclosed company[80].

2. **Company valuation methods based on other indicators and criteria.** There are other less commonly used valuation methods for quantifying harm, such as methods based on the balance sheet (book value, liquidation value), on the income statement (sales, EBITDA), on goodwill (calculation of the value of a company based on its brand value), value creation (economic profit taking into account the evolution of equity), and other discounted flow methods (in addition to NPV,

[79] A more detailed discussion of some of the main advantages and disadvantages of using margins can be found in Oxera (2015).

[80] Another method frequently used to measure profitability is the internal rate of return (IRR), which indicates the interest rate that would make the NPV of the project or company be valued at zero.

dividend discounting methods are often used to calculate the trend in the value of a company's shares based on the future dividend payout).

3. The **cost of capital**. The lower limit of the loss of profit can be quantified by calculating the cost of capital, i.e., the cost incurred by the company to finance its investment projects through both equity and borrowed funds. The cost of capital can provide an estimate of the minimum profit margin a company needs to remain profitable to investors (and stay in business), although this will depend on the circumstances of the case[81].

111. Once the company's profitability has been calculated, the second stage involves defining a counterfactual to estimate what the profit would have been in the absence of the infringement, using any of the comparative methods already described (e.g., profitability before and after the infringement, or comparing profitability over time with another company with a similar market and characteristics). The main **advantage** of financial methods lies in the accessibility and reliability of certain accounting and financial data. This derives from legal disclosure and auditing obligations, which are greater for listed companies.

112. As for the **drawbacks** of these methods, it should be noted that isolating the impact of the infringement on financial performance from other relevant factors can be challenging, and similar caveats apply as for comparative methods. Difficulties may also arise when defining profitability not only for the counterfactual but also for the actual scenario, and the variables (e.g., cash flows) chosen may need to be justified in detail, depending on the approach and the particularities of the company and sector under analysis.

Box 1 Business profits and margins

As highlighted above, determining the cost overcharge plays a fundamental role in quantifying harm. In the framework of the cost method, primarily relying on company accounting data, it is **necessary to distinguish between several** related **but distinct concepts**:

[81] For example, if the counterfactual market structure is characterised by imperfect competition (entry barriers, low number of competitors, etc.), the companies in the sector may maintain a profitability above the cost of capital for an extended period. Even in competitive markets, profitability can differ from the equilibrium in specific periods.

Gross profit: This is the result of subtracting costs directly related to sales from total sales for a specific point in time. Gross profit is a measure of the ability to obtain results directly linked to a particular activity.

Operating profit: This is the result of subtracting operating expenses from gross profit. When only operating expenses are deducted, this is what is known in accounting terms as earnings before interest, taxes, depreciation, and amortization (**EBITDA**). This indicator provides information on the company's capacity to generate resources through its ordinary activity.

When, in addition, depreciation and amortisation of the company's assets are deducted, the operating profit is known as earnings before interest and taxes (**EBIT**). This indicator allows companies in the same sector to be compared from a purely operational standpoint, as it excludes financial and fiscal factors, the nature of which may be heterogeneous.

Net profit: This is the result of deducting all remaining expenses (mainly financial and tax expenses) from EBIT.

Business margin: Building upon the concept of profit, the main difference lies in the nature of the indicator. While profit is an absolute measure (an amount expressed in euros), the business margin is a relative measure, expressed as a ratio of profit (gross, operating or net, depending on whether we are calculating gross, operating or net margin, respectively) to revenue (sales). Thus, the business margin shows, as a percentage, how much of each euro of revenue is translated into profit. This facilitates comparisons with other companies in the same sector that are of different sizes.

How are business margins obtained?

As emphasized, quantification of harm focuses on trying to recreate the scenario in the absence of the anticompetitive conduct in question. Applying the cost-based method, the objective is to compare the **business margin** resulting from the infringement with the **business margin deemed reasonable that would have occurred in its absence**.

To facilitate the understanding of these financial concepts, a simplified example of a profit and loss account is shown below:

No.	Concept	Amount (millions)
1	Sales	36,772 €
2	Merchandise cost	14,975 €
3	Gross Profit (1) - (2)	21,797 €
4	Gross margin (3) / (1)	59.3%
5	Operating expenses	9,811 €
6	EBITDA (3) - (5)	11,986 €
7	Operating margin (6) / (1)	33%
8	Amortizations and depreciations	3,391 €
9	EBIT (6) - (8)	8,595 €
10	Operating margin (9) / (1)	23%
11	Financial expenses (interest)	182 €
12	Results before taxes (9) - (11)	8,413 €
13	Profit taxes	1,241 €
14	Net profit (12) - (13)	7,172 €
15	Net margin (14) / (1)	20%

Source: prepared in-house.

2.3.3. Simulation models

113. **Simulation models** are based on **economic theory** (industrial organisation models and game theory) and include **data** to simulate and try to predict the **behaviour of the agents in the market** in the absence of an infringement.

114. These models aim to **simulate the value of the variable of interest** (e.g., price, business margin, market share or level of output), defining in advance the most appropriate characteristics of the market to be simulated within the model. The key elements to be defined are mainly the **characteristics of supply** (type of competition between firms, degree of market concentration, barriers to entry, product differentiation, cost structure, etc.) **and demand** (especially price elasticities and cross-price elasticities). To achieve this, it is necessary to construct a system of equations, with several variables and parameters. The objective is to simulate the equilibrium that would be reached in the market on the defined supply and demand characteristics. The parameter values can be known, econometrically estimated, or assumed (depending on the complexity of the model and data availability) ensuring that the model is consistent with the main characteristics of the market.

115. These models enable the construction of different scenarios tailored to the specific circumstances of each individual case and analyse the impact of anticompetitive behaviour on the outcomes of the companies of interest.

Although the models will necessarily simplify reality, they must reflect the main features of the competitive interaction between market agents[82].

116. Two main approaches can be distinguished when using simulation models to quantify harm (Oxera, 2009). These approaches share the need to select the best model for describing the behaviour of the market in the counterfactual scenario:

1. One option is to develop a **single model** to represent competition in the counterfactual scenario[83]. To calibrate this model, market demand and relevant characteristics on the supply side are estimated. In most cases, this estimate is obtained with econometric methods. The resulting model is used to obtain the simulated results of the counterfactual and then compare them with the actual data and quantify the harm.

2. Another option is to develop **two theoretical models**, one for the counterfactual scenario and one for the observed scenario. Data from the observed scenario (prices, quantities, costs) can be used to infer market demand characteristics and simulate the counterfactual, without the need for econometric estimation. This approach may be technically

[82] The main models used by economic theory to represent the interaction of companies in a market are:

- Perfect competition: a high number of sellers and buyers, homogeneous product, absence of barriers to entry and exit, perfect information and agents without the capacity to individually influence the market price, which is equal to the marginal cost.
- Monopolistic competition: a large number of companies with differentiated, but similar substitute products and reduced barriers to entry. Differentiation provides each firm with some market power, which allows them to raise the price above the marginal cost.
- Oligopoly: the existence of a reduced number of companies among which there is strategic interdependence (the outcome of each company's strategy will be affected by the strategy followed by its competitors). The Cournot and Bertrand models are the most common, depending on whether the companies compete in terms of quantity or price, respectively. While the Cournot model reaches a situation of intermediate prices and quantities between perfect competition and monopoly, Bertrand's model arrives at the same equilibrium as seen in perfect competition (price equals marginal cost). If any of Bertrand's suppositions are lifted and features such as product differentiation or capacity restrictions are introduced, the result is further from perfect competition. There are also dynamic oligopoly models such as the Stackelberg (a leading company makes the first move, and the other companies respond to this action) and Dixit (this models the decision to enter the market in the presence of strategic barriers) models. In recent years, economic models of bilateral bargaining with strategic interactions have also emerged as an alternative ("*bargaining models*").
- Monopoly: a single company produces a good and sets the market price (equalising marginal revenue with marginal cost). The only limitation is market demand, which establishes the quantity sold at the chosen price (or, alternatively, determines the market price given the quantity produced by the monopolist). The price will be higher, and the amount exchanged lower than any other model.
- Auction models: these are often used to represent anticompetitive practices in auction markets. For example, these include *bid rigging* practices, which involves agreements between competitors in public tenders to fix the prices and/or other conditions of the bids presented.

[83] Real data within the scope of the infringement will typically be used, although assumptions will also have to be made about the characteristics of the counterfactual, when it is considered that the infringement has modified the structure of the market.

less complex and require less data. However, to obtain valid results, the assumptions underlying the models used to simulate the counterfactual and observed scenarios become more critical.

117. One of the main **challenges** of these methods is to model the type of competition that would have existed in the factual and, in particular, counterfactual scenarios. It is crucial to justify the competition model used to describe the interaction between market players, as this choice may lead to significantly different outcomes. This is because demand and supply affect price formation differently in the case of a monopoly structure than in the case of a competitive market (i.e., a market with perfect or imperfect competition, or oligopolistic competition), and this will substantially affect the final outcome of the harm quantification[84].

118. One of the **advantages** of these models is that the results are derived from the rational behaviour of companies, combined with observed data on the characteristics of demand and supply in the analysed market. In addition, these methods allow the analysis to incorporate, among other factors, changes in market structure that may have been caused by infringements. Furthermore, where no suitable counterfactual exists due to the particularities of the case (e.g., due to lack of data), a valid hypothetical infringement-free scenario can be constructed using simulation models.

119. The **drawbacks** of these methods include its theoretical and factual requirements, and the need for detailed information on the form of competition between companies and the determinants of supply and demand[85]. Moreover, the results are likely to vary substantially depending on the assumptions adopted, making it essential to demonstrate the robustness of the model.

[84] Sometimes it is assumed that the companies participating in a cartel behave as if they were a single monopolistic entity, coordinating to maximise their joint profits. Although one might expect this to be the conduct of a fully effective cartel, joint benefits are typically lower because of the incentives for firms to deviate from collusion, which increases with the number of firms, their asymmetry, and the opacity of the market, among other factors. Therefore, the monopoly situation could be used as the upper limit of the overcharge reached in a market with infringement, making it necessary to analyse the details of each case.

[85] This complexity will be greater the more econometric estimation of the parameters is used, while the information requirements will be less if calibration techniques are used.

2.4. Capitalisation of the harm

120. As mentioned above, the LDC stipulates that **full compensation** for an infringement of competition law "shall comprise compensation for actual loss and loss of profit, plus interest"[86].

121. Once the harm suffered by the plaintiff at the time of the infringement has been assessed, it is necessary to capitalise it, i.e., express it in current terms, by applying a capitalisation or interest rate. In this way, the aim is to take into account the "time value of money", that is, to reflect the fact that a monetary amount has a different value depending on when it materialises. This is because, over time, various factors can affect the value of money, such as, among other things, inflation and the return on investments made.

122. Capitalisation of the harm is therefore a fundamental issue, which must be taken into account both by the parties and by the courts, as it can represent a considerable fraction of the final compensation (especially in more protracted infringements). This is an area with a dual economic and legal aspect, insofar as it interrelates legal obligations (sometimes the law or case law may stipulate the amount or the way in which interest is calculated) with economic principles, notably concerning the preference of applying a particular capitalisation rate depending on the circumstances of each case.

123. In legal proceedings, it is most common to capitalise the harm, since harm resulting from anticompetitive conduct usually predates the date on which the quantification is made. However, it may sometimes be necessary to perform the reverse operation and express a future monetary amount in present value, which is known as "discounting" or " updating"[87]. Capitalising the value of past harm and discounting the value of future harm is necessary to be able to quantify possible harm occurring at different points in time and express it on the same comparable basis.

---

[86] Directive 2014/104/EU (considering 12): *"[…] The payment of interest is an essential element of compensation to repair the harm suffered, taking into account the passage of time, and must be required from the moment the harm occurred until the moment the compensation is paid, notwithstanding that under national law such interest is classified as compensatory interest or late payment interest, and that the passage of time is taken into account as an independent category (interest) or as a constitutive part of the loss suffered or loss of profits. It is up to the Member States to establish the rules that must be applied for this purpose."*

[87] For example, anticompetitive conduct may have lasting effects in the medium/long term, so that part of the harm claimed at the time of preparing the expert report corresponds to a future period. In that case, future harm would have to be discounted to obtain its current value.

2.4.1. Methods of calculating capitalisation

124. The capitalisation of harm can be carried out according to two methods: **simple** or **compound capitalisation**.

125. When applying the simple capitalisation method, the value ultimately obtained will be equal to the initial capital plus the interest generated in each period. This interest is calculated by applying the capitalisation rate only to the initial capital.

126. In contrast, when applying the compound capitalisation method, the interest generated in each period will depend on (i) the initial capital and (ii) the interest generated in all previous periods. Thus, in each period, the capitalisation rate will be applied to the sum of the initial capital and the interest accrued up to that time. From an economic perspective, **the compound method is the most comprehensive and usually recommended, as it takes into account the fact that interest can be reinvested as it is earned**.

127. To give a simple example, we will assume that:

   i) a one-off infringement over time gives rise to a harm quantification amounting to 1,000 euros, measured in euros at the time the harm occurred,

   ii) the applicable annual interest is 10%, and

   iii) three years have passed between the time the harm occurred and the time the lawsuit is filed.

128. In this case, with simple capitalisation, the interest would amount to 300 euros, while with compound capitalisation, the total interest would be 331 euros, as detailed in the following table:

Table 1. Example for comparing the results of simple and compound capitalisation[88]

	Simple capitalisation $V_{final} = V_{initial} \cdot (1 + n * i)$	**Compound capitalisation** $V_{final} = V_{initial} \cdot (1 + i)^n$
Initial value	1,000€	1,000€
Annual cap. rate (i)	10% (0.1)	10% (0.1)
No. years (n)	3	3
3rd year cap.	$1{,}000€ * (1 + 3 * 0.1) = 1{,}300€$	$1{,}000€ * (1 + 0.1)^3 = 1{,}331€$
Total interest	$1{,}300€ - 1{,}000€ = 300€$	$1{,}331€ - 1{,}000€ = 331€$

Source: prepared in-house.

2.4.2. Capitalisation rates

129. Another fundamental question that must be determined is the capitalisation rate or interest rate that should be applied in each case. The higher the rate applied for the capitalisation of past harm, the greater the harm expressed in present terms[89]. Although there is no clear consensus in the literature, the following are some of the rates that could be used[90]:

1. The legal interest **rate** is applied, following the Civil Code, as compensation for harm when the debtor defaults on debt payments and there is no other rate agreed upon by the parties. It is fixed for each year by the General State Budget Law [Ley de Presupuestos Generales del Estado]. Although it is quite commonly used to capitalise harm, it does have disadvantages such as its lack of adaptability to changes in market conditions (due to its low periodicity) or to the particular circumstances of each claim.

2. Risk-free interest rate This is usually calculated based on of the return on government debt instruments (bills, bonds, debentures), as these are generally low-risk financial instruments. This interest rate could be taken as the minimum capitalisation rate, since the risk taken by private operators in their investments is usually positive[91]. A risk-free rate can also be used, if it is considered that the plaintiff company must accept

[88] $V_{initial}$ is the monetary amount to be capitalised, V_{final} is the amount resulting from applying the method (initial amount + interest), *i* is the interest rate applied (for simplicity, we assume that it is the same in all periods, although it normally varies over time) and *n* is the number of periods between the two points in time (initial and final).

[89] The opposite will be true if future harm is discounted: the higher the discount rate, the lower the present value.

[90] For more information on capitalisation rates, see Oxera (2006), Gotanda and Sénéchal (2009), Compass Lexecon (2017) and Dow (2022).

[91] Other reference interest rates such as the interbank interest rate (e.g., Euribor or €STR in the Eurozone) could also be used.

without compensation all the defendant's insolvency risk during the legal process.

3. The **cost of the defendant's debt**. The defendant could be considered to have, in a sense, taken a forced loan from the plaintiff by causing the harm claimed. The plaintiff should therefore be compensated for bearing the risk of the defendant's insolvency from the time of the harm until the date of the ruling of the first instance. This could be estimated from the cost of debt issued by the defendant.

4. The **plaintiff's cost of capital**. The weighted average cost of capital (WACC), which attempts to estimate the return demanded of the plaintiff company by its investors[92], is normally used to reflect the opportunity cost to the plaintiff of the unavailability of economic resources between the time the harm occurred and the compensation settlement. This method therefore assumes that the plaintiff has to be compensated for the cost of the additional capital needed to cover the reduced income or increased costs resulting from the initial harm.

130. The **Consumer Price Index (CPI)** is also sometimes used to compensate for monetary depreciation caused by the passage of time. However, the CPI, unlike other rates, only takes into account inflation and does not reflect the cost of the plaintiff's lost opportunities for not having had that capital in the past. Therefore, from an economic point of view, its use as a capitalisation rate is not recommended. Additionally, it will be necessary to consider whether the quantification method used has led to results expressed in nominal or real terms, since, in the latter case, inflation would not have to be taken into account again when calculating the capitalisation of the harm.

2.4.3. Temporal delimitation of interests

131. In general terms, when calculating interest in the context of the quantification of harm, three periods can be distinguished, which we will call A, B and C:

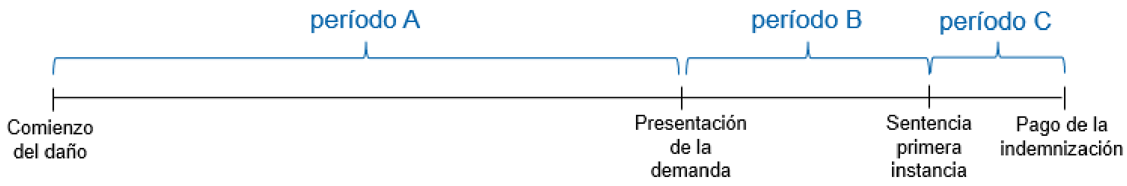
- Period A: from the time the harm occurs until the date the claim is filed.

92 The WACC (*weighted average cost of capital*) is the average cost of the two capital resources that a company has (debt and equity), weighted by their relative weights in total liabilities. Another possibility is to use only the cost of one of the two types of financing as the capitalisation rate:

- The cost of own funds, in the particular case that the plaintiff is a shareholder.
- The cost of the debt, if the plaintiff had to go into debt because they did not have the resources they would have had in the absence of the infringement.

- Period B: from the filing of the application to the ruling of the first instance.
- Period C: from the ruling of the first instance until the compensation is received.

Figure 2. Periods for calculating the interest



Source: prepared in-house.

132. The interest in periods A and B would be compensatory in nature and would be justified by the principle of indemnity, which advocates the right to complete reparation of the harm caused[93]. In both cases, they form part of the claim, so they must be expressly requested in the claim so that they can be granted. Although it is common to use the legal interest rate to capitalise the harm, in principle, nothing excludes the application of other alternative capitalisation rates, when it is considered that these allow greater compliance with the principle of indemnity[94].

133. The interest for period C is the interest on the procedural delay (interest in arrears) referred to in Article 576 of the LEC, intended to compensate the creditor for the time elapsed from the moment the judgement is issued, even if it is not final, until the compensation is received. The difference with respect to periods A and B is that, in period C, the applicable interest is determined in accordance with the provisions of the LEC (Article 576.1)[95]

[93] The Supreme Court makes this clear in its judgements: STS 2472/2023, , of 12 June 2023 - ECLI:ES:TS:2023:2472; STS 2473/2023, of 12 June 2023 - ECLI:ES:TS:2023:2473; STS 2475/2023, of 12 June 2023 - ECLI:ES:TS:2023:2475; STS 2476/2023, of 12 June 2023 - ECLI:ES:TS:2023:2476; STS 2477/2023, of 12 June 2023 - ECLI:ES:TS:2023:2477; STS 2479/2023, of 12 June 2023 - ECLI:ES:TS:2023:2479; STS 2480/2023, of 12 June 2023 - ECLI:ES:TS:2023:2480; STS 2492/2023, of 12 June 2023 - ECLI:ES:TS:2023:2492; STS 2494/2023, of 12 June 2023 - ECLI:ES:TS:2023:2494; STS 2495/2023, of 12 June 2023 - ECLI:ES:TS:2023:2495; STS 2497/2023, of 12 June 2023 - ECLI:ES:TS:2023:2497.

[94] See, for example, STS 123/2015, March 4, 2015, ECLI:ES:TS:2015:669.

[95] It is the interest for procedural default, i.e., the legal interest plus two points or that which corresponds by agreement of the parties or by special provision of the law.

and is applied *ex officio* by the court, without the need for it to be requested in the claim[96].

134. It should be borne in mind that it is relatively common for anticompetitive conduct to last longer than one year and even for different types of damages to be claimed. In such cases, the capitalisation of the harm for period A could be divided into several steps (which should be clearly and transparently stated in an expert report):

i) For each period, the different types of harm are added up separately in nominal terms.

ii) The harm for each period is appropriately capitalised (see example above under Subsection 2.4.1). This will allow the monetary amounts of the different periods to be comparable to each other.

iii) The harm for the different periods is added together to obtain the total amount claimed at the time of filing the claim.

135. Regarding period B, the calculation cannot be made when filing the claim because the time until the ruling of the first instance is unknown, but the claim should explicitly state the request for interest payments, the proposed interest rate and the calculation method to be applied.

2.5. Differences in the quantification of harm caused by price increases and exclusionary practices

136. In recent years, most of the theoretical discussions and legal proceedings on damages claims have focused on cases of price increases, specifically, on cartel behaviour. Although the discussion in the previous sections of the Guide has tried to not limit itself to the harm arising from a specific conduct, it is considered necessary to delve on the **particularities of exclusionary conduct** in this section.

137. First, the actors involved and the type of harm are different. Conducts leading to overcharging mainly affect buyers (vertical dimension), who usually claim actual loss[97]. In the case of exclusionary practices, the directly affected actors are usually competitors (horizontal dimension), **who suffer harm due to lost sales resulting in lower profits (loss of profit)**. However, buyers may also be affected both negatively (higher prices and lower quality and variety after the exclusion has ended) and positively (e.g.,

---

[96] Monti (2016), pp. 271 - 289, or Ruiz Peris et al. (2021), pp. 275 - 297.

[97] If there is a pass-on effect of costs and volume, loss of profit could also be requested, although it is not common.

through discounts or predatory pricing in the initial phase of the conduct). Similarly, suppliers of the foreclosed firms may also be affected.

138. Secondly, the harm associated with price increases is typically more direct: price increases and quantity decreases compared to the no-infringement scenario. Moreover, the market structure generally remains unchanged. In contrast, **foreclosure cases tend to generate more complex effects** that vary over time and may alter the market structure, making it difficult to return to the situation that existed prior to the infringement.

139. Third, the **problems of missing data** are often greater in exclusionary practice damage claims. While practices that lead to price increases do not usually prevent injured parties from continuing to purchase the affected products, exclusionary practices may permanently exclude certain competitors from the market (leading to a lack of post-infringement data) or prevent potential competitors from entering the market (leading to even greater data limitations). This makes the construction of the infringement counterfactual scenario more difficult, which, in turn, makes the quantitative study of the effects of foreclosure more complex.

140. The following is a general explanation of the main effects of exclusionary practices over time, followed by a discussion of the main tools that can be used to quantify them[98].

2.5.1. Particularities of exclusionary practices

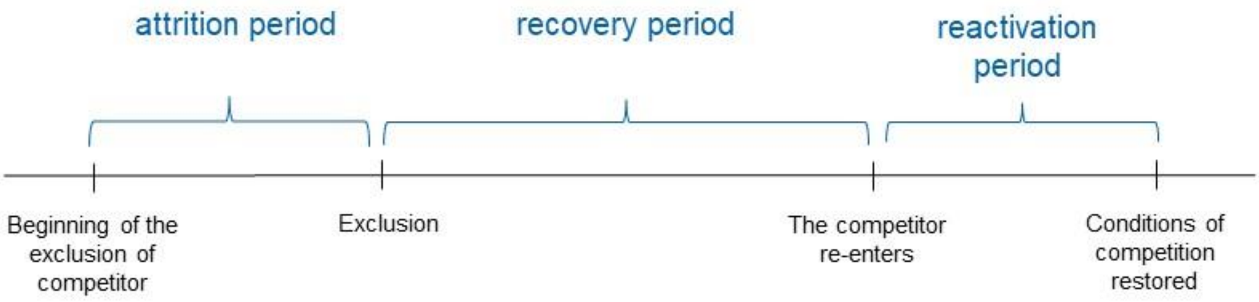
141. This subsection describes, from a theoretical point of view, the effects of exclusionary conduct. This category encompasses a wide range of practices, each with its own peculiarities, but an attempt will be made to give an overview of the most common types of harm. In the **timeframe** of exclusionary practices, three distinct phases can be separated between the moment of the adoption of the anticompetitive practice and the possible restoration of competitive conditions in the market. These three phases are the attrition period, recovery period, and reactivation period[99].

---

98 For a more detailed analysis, see García et al. (2018), Fumagalli et al. (2010), Buccirossi (2010) and Prosperetti (2009).

99 Fumagalli et al. (2010) and García et al. (2018).

Figure 3. Phases of exclusionary behaviour



Source: prepared in-house based on Fumagalli et al. (2010).

142. The **attrition period** starts with the adoption of the exclusionary strategy and ends once the reduction of the rival's market share or its temporary or permanent foreclosure has been achieved. During this stage, there is aggressive competition by the dominant company to reduce the market share of its competitors, which may lead to defensive strategies by the latter, involving significant sunk costs and investments. Finally, the reduction in market share implies a reduction in revenue which, in turn, may lead to higher unit costs and a loss of profits, thus reducing the possibility of remaining in the market.

143. Another type of harmful behaviour is the creation of **barriers to entry** that prevent potential competitors from making profits in this market. Moreover, **price wars** (e.g., predatory pricing) are frequently observed at this stage, which negatively affects the profit of firms competing at the same level and favours, at least temporarily, buyers. However, if the conduct seeks to deteriorate the position of a "downstream" competitor by increasing its costs or refusing to supply it, buyers may be adversely affected.

144. In the **recovery period**, the infringer's market power is greater, as its competitor(s) has withdrawn from the market, failed to enter or has lost market share. During this period, the infringer will be able to make use of its **dominant position** (strengthened after the conduct), which will allow it to increase prices and recover the profit lost in the previous period. The quantification of harm has to consider two different aspects. On the one hand, there could be horizontal harm in the form of unrealised profits (loss of profit) for any excluded competitors. On the other hand, there could be a vertical effect, harming direct and indirect purchasers if the price increase materialises.

145. Finally, the **reactivation period** starts after the end of the abusive conduct (in follow-on cases, it may coincide with the administrative decision that establishes the cessation of the conduct), leading to the progressive re-

establishment of competitive conditions: the same or new companies enter or gain market share. The effect on market prices is ambiguous and can be a positive or negative change, depending on the competitors' strategy.

146. These three periods describe, in a simplified way, **a complete cycle of how exclusionary conduct evolves**. However, the first does not necessarily have to occur, especially when, in vertically connected markets, the incumbent company signs exclusive agreements, thereby leading to input exclusion or consumer exclusion and preventing potential competitors from entering[100]. Moreover, in markets with high barriers to entry (due to, among other things, the existence of significant economies of scale, scope, or network economies), the recovery period may be protracted, or the start of the upswing (reactivation phase) may not even begin. It should also be borne in mind that the timing and manner of competition authorities' intervention may affect the duration of the different stages or their absence.

2.5.2. Relevant issues when quantifying harm

147. In cases of exclusion, the loss of profit concept becomes more important (which does not preclude the existence of actual loss). This is particularly true for competitors who are excluded, as they may see their costs increase, prices decrease, or sales volume decrease due to the exclusionary conduct[101]. To calculate this, it is necessary to compare the profits made in the affected markets during the time the infringement has had an effect with those that would have been made in the absence of the infringement.

148. As in any other damages claim, a fundamental element is the proper construction of the counterfactual scenario. The absence of anticompetitive conduct does not imply that the economic activity is conducted in a perfectly competitive market, but rather that different market typologies are possible. Therefore, knowledge of the characteristics of the affected market is necessary so that the quantification is as precise as possible. In follow-on complaints, the administrative decision and possible commitments or conditions imposed may be useful.

149. Next, the effects on excluded competitors are discussed. Regardless of the phase analysed, it is important to have detailed data on prices, costs and sales volumes. In the attrition period, exclusionary conducts may lead to **increased costs and reduced revenues** for competing firms. This may lead to a reduction in market share and ultimately to incur costs when exiting

[100] According to the OECD (2011), these are the two most common forms of vertical anticompetitive conduct.

[101] These effects may be interrelated: the increase in costs can make the business activity unfeasible, while a drop in sales can increase the unit cost.

the market. When quantifying the harm, it is necessary to distinguish between the exit costs stemming from the infringement and the sunk costs that the firm would have incurred.

150. Competitors' **loss of profits** tends to concentrate in later periods as long as effective competition is not re-established in the market. While in the quantification of actual loss prices are of fundamental importance, in cases of loss of profit other variables such as sales volumes, market shares or business margins gain weight to approximate lost profits.

151. When constructing the counterfactual, the techniques mentioned in this Guide can be used. One possibility would be to use comparative methods, taking as a benchmark the results of the damaged company in a period unaffected by the conduct or in a different market, as well as the performance of similar companies. The greater difficulty in finding time periods or comparison markets in exclusion cases may favour the use of simulation models, which try to approximate the situation the competitor would have been in the market in the absence of the infringement. Financial methods may also be useful for approximating the counterfactual profitability of the excluded company and quantifying the loss of profit.

152. It should be borne in mind that, given the hypothetical behaviour of other market players, it may be more difficult to obtain data to quantify the loss of profit from exclusionary conduct than to quantify the direct effect of an overcharge. It is, therefore, necessary to make predictions about the hypothetical (often future) profits of a particular company in the market, so that the factors that may affect the estimation are multiplied and, consequently, the precision of the estimates tends to decrease. This makes it necessary to consider the specificities of the companies concerned and to adjust the results of the selected methods based on often qualitative information.

153. For example, in the case of a firm that was excluded before it started operating, in the total absence of information on its performance in the affected market, factors such as its technological level relative to the dominant firm could be considered to predict how its market share would have evolved. In the absence of additional information, one could assume that the competitor would have had similar technology to the dominant company[102]. On the other hand, if the plaintiff company was active in the market before being excluded, its pre-infringement performance may be

102 This approach is in line with the *As-Efficient-Competitor Test*, which is sometimes used as a reference to test the effects of abuses of dominance.

useful for approximating how its position would have evolved in the absence of infringement.

154. Another possibility often used by excluded competitors is to seek reparation only for the additional costs incurred for the business activity that was ultimately frustrated by the infringement[103]. In this way, it is not necessary to make assumptions about foregone profits, although the harm actually incurred would be underestimated (this could be a way of calculating a lower bound for the harm).

155. The effect on consumers may be diverse, ranging from price increases (in which case, similar considerations would apply to the quantification of overcharge harm), to loss of product quality or variety. Additionally, potential positive effects that have occurred in the attrition period in the form of lower prices or other efficiencies resulting from the conduct should also be considered.

2.6. Checklist for testing quantification reliability

156. The following are indicative, non-exhaustive and not mutually exclusive **methodological caveats** that could be used to check how reliable the results of harm quantifications are.

2.6.1. General checklist

How has the market affected by the infringement been described?

157. Analyse whether the expert report reflects the main characteristics of the affected market, considering aspects such as the structure and maturity of the markets, the degree of competition, determinants of supply and demand and any other issues that may affect the quantification.

158. In cases of *follow-up* claims, analyse whether the expert reports take as a starting point the elements identified in the final decision, justifying any divergence from this.

Has the theory of harm been adequately described for the specific case?

159. Analyse whether the expert report reflects the type of infringement and harm suffered (actual loss or loss of profit) and the mechanism through which the plaintiff has been harmed.

[103] This practice is less demanding in terms of data collection and the construction of counterfactuals, although from a conceptual perspective it could be less complete.

160. Assess whether the theory of harm described in the expert report provides a sound economic explanation and is consistent with the applicable legal presumptions (e.g., Art. 76.3. of the LDC for cartels) depending on the characteristics of the actors involved, the types of infringement and the type of claim (stand-alone or follow-on).

What assumptions and hypotheses have been considered in the construction of the counterfactual?

161. Analyse whether the expert report explains the criteria used to construct the counterfactual in a detailed and transparent manner (this in turn must be compatible with competition law). Also, whether factors that may lead to an over- or underestimation of the actual harm are included, as well as the reasons for rejecting other possibilities for the counterfactual.

162. Analyse whether it explains why the assumptions and hypotheses used can be considered reasonable and likely and allow for the construction of a scenario that is a reasonable approximation of what would have happened in the absence of the infringement (counterfactual). For this purpose, techniques can be used to assess the similarity between the two scenarios (e.g., testing the means, parallel trends, correlations of prices or other variables, etc.)[104].

What has been considered to select the variables?

163. Analyse whether the expert report contains a detailed justification of the variables chosen to quantify the harm, based on the theory of harm and economic theory (so as not to omit variables that according to economic theory could affect the market in question or, conversely, to include irrelevant variables that could lead to spurious correlations and biases and inconsistencies in the quantification)[105].

164. Examine whether, once the variables have been selected, the expert report includes a descriptive analysis of these variables (definition, description, evolution, relationship between them and with the variable to be explained, justification for their selection), prior to the use of the quantification method.

165. Analyse whether the expert report also addresses how the effects caused by other factors unrelated to the infringement have been isolated, if they exist.

---

[104] See Annexes 2 and 4.

[105] See Section A2.2.2 of Annex 2.

Has the temporal delimitation of the infringement been adequately reasoned?

166. Examine whether the duration of the effects of the infringement has been reasonably and transparently delimited. In follow-on actions, any deviations from the period included in the competition authority's decision should be justified. If there are doubts about the inclusion or not of certain periods, the justification should be more comprehensive, and the use of quantitative techniques being recommended whenever the data allows this.

How has the database been designed?

167. Analyse whether the expert report describes in detail (i) the data collection process; (ii) the characteristics of the sample and its degree of representativeness; and (iii) the processing of the original database, including the detailed justification for any modification thereof.

168. Examine whether the expert report explains the difficulties encountered in accessing the necessary data and specifies whether it finally deems the database to be sufficiently complete to perform the quantification.

How were the selected quantification method(s) chosen and applied?

169. Justify the selection of quantification methods and techniques used and analyse whether they have been applied following best economic practices. No approach should be discarded *a priori*; instead, it is recommendable to examine whether the necessary methodological care (described in the following section) has been applied and a sufficient level of precision achieved, considering the particularities of each case and the constraints imposed by the information, resources and time available.

170. Examine whether the parties have conducted their analyses as transparently as possible in relation to the data used, assumptions and limitations of the models, among other issues.

When using econometric techniques, have the appropriate methodological care and tests been applied?

171. If econometric techniques are used, analyse whether the expert report justifies the specification of the model and whether it presents problems of, among other things, endogeneity, heteroscedasticity, autocorrelation or multicollinearity[106].

---

[106] See Annexes 2 and 4.

172. Also, examine whether the expert report analyses the estimated coefficients of the variables of the model (especially those quantifying the effects of the infringement) from the perspective of their sign, magnitude and statistical significance.

173. Likewise, it should be analysed whether there is a reasoned explanation for the absence of bias in the coefficients of interest or, if bias exists, its implications should be discussed.

174. Without prejudice to confidentiality considerations, it would be good practice to make available the data sets and codes used in the quantitative estimations to promote the reproducibility of the results.

Has the overcharge pass-on been conclusively analysed?

175. Examine whether the expert report sufficiently analyses the existence of overcharge pass-on. It is important to know: (i) who has alleged it (the defendant, the plaintiff or both); (ii) whether it has been sufficiently proven, taking into account the standard of proof and who bears the burden of proof; (iii) whether the particular characteristics of the market analysed have been taken into account and how they affect the pass-on rate; (iv) whether the measurement of the volume effect is also included.

Is there a complete presentation and evaluation of the results?

176. Examine whether the expert report includes a sensitivity analysis of the results to check how the specification, assumptions or set of variables included influence their determination. In this way, the analysis and its results will be more robust.

177. Analyse whether it has been sufficiently reasoned that the result obtained is a minimum or maximum estimate of the harm. It is also advisable to specify confidence intervals to capture the degree of uncertainty around the estimated harm or to use appropriately explained graphs.

178. Check that methods and conclusions are presented in a straightforward and accessible way, making them easy to understand.

Has the harm been correctly capitalised?

179. Analyse whether the harm calculated is properly capitalised: in a first stage, from the time it occurs until the claim is filed; in a second stage, between the filing of the claim and the date of the ruling of the first instance; and finally, in the subsequent period, until there is effective compensation.

180. It is necessary to explicitly claim the capitalisation of the corresponding harm in the first periods referred to, indicating in detail in the expert report

the rate of capitalisation and the calculation method (simple or compound) to be applied.

2.6.2. Specific checklists

181. In addition to the general methodological precautions, it is advisable to take into account a set of indicative and non-exhaustive checks when applying the different quantification methods.

Comparative methods

182. Analyse whether it has been justified that the markets used as comparators have not been influenced by the infringement and are sufficiently similar to the affected market.

183. Check that the comparison period is completely separated from the effects of the infringement (by applying the parallel trends assumption in the case of the difference-in-differences method).

184. Check that the comparability of the observations of the scenarios with and without the infringement has been enhanced by using statistical tests, qualitative or factual data on the comparison groups.

185. Analyse whether, in studying the impact of the infringement on the variable of interest, other factors that may have affected the variable of interest have been taken into account.

186. Verify that, if the evolution of the variable of interest has a strong seasonal component, an attempt has been made to isolate the impact of these periodic effects.

187. If temporal comparisons of monetary variables (prices, business margins, costs) are conducted, examine how the effects of inflation and exchange rates have been taken into account, where applicable.

Cost-based methods and financial analysis

188. Analyse whether the chosen cost characteristics have been justified in terms of the specific circumstances of each case, whether consistency has been maintained concerning the margins considered, and whether there is an explanation of issues such as:

- the types of costs taken into account and how the unit cost was calculated,

- how costs common to other products or services have been apportioned,

- whether any adjustments have been made to the accounting data to bring it closer to the economic concepts,

- if the observed costs differ from those that would have occurred in the absence of infringement.

189. Examine that the references used to obtain the "reasonable" margin applied to the cost or profitability of the company under analysis have been explained, and that they have been justified as valid approximations of its situation, both actual and counterfactual. To this end, it should be verified that the main external factors to the infringement that may affect the analysis (e.g., sector characteristics) have been taken into account. In this respect, similar considerations apply as for comparative methods.

Simulation models

190. Verify that the supply side of the model has been adequately justified. In particular, the following points should be analysed:

- The selected competition model fits the market characteristics, performance and observed behaviour of the companies.

- The extent to which the modelled cost structure reflects the reality of the companies in the sector.

- The existence or absence of capacity constraints, if relevant.

- Whether the products concerned are homogeneous or differentiated.

191. Analise whether the demand side of the model has been adequately justified, especially if it is decided to estimate the function. In particular, that decisions have been justified on, *inter alia* and depending on the model used, any of the following points:

- The selection of the demand function used.

- How own and cross-price elasticity as well as income elasticity have been considered.

- The inclusion of prices or quantities of substitute or complementary products.

- The inclusion of socio-economic data determining the level of demand (e.g., consumer income).

- Market concentration.

192. Verify that it has been demonstrated that the model reasonably explains the counterfactual scenario. This may require justifying that the model partially fits the actual market data and that the remaining assumptions are consistent with what would be expected in the absence of infringement.

3. CONCLUSIONS

193. **Private enforcement of competition law is of paramount importance** because it contributes to ensuring that the positive effects of effective competition reach all economic agents. It allows victims to be compensated for the harm suffered and deters operators from engaging in anticompetitive infringements. In this sense, it complements the public enforcement of competition law.

194. However, **quantifying harm can present significant difficulties in some cases**. The main issue is that it requires a comparison of the actual economic situation of the injured party with a reasonable approximation of the **hypothetical (counterfactual) situation** in which they would have been had the anticompetitive conduct not occurred. To carry out this task, **multiple quantitative-oriented methodologies** can be used, based on disciplines such as microeconomic theory, econometrics, corporate finance, and industrial organisation. **The choice of the most appropriate methodology depends on each specific case**: type of infringement, available data, level of evidence required and proportionality between the associated costs (resources, time), among other things.

195. This Guide has a **purely advisory** value, since the settlement, quantification or estimation of harm is the responsibility of the competent judicial body. The Guide is intended to **provide information to all those involved in the calculation of harm (judges, courts and operators)** on criteria and aspects to be taken into account so that they can determine which methods are most reliable and appropriate for quantifying harm in each specific case. At the same time, the Guide is aimed at **facilitating the exercise of claims for damages** and at **disseminating best practices** when quantifying harm, thus being useful for all parties to the proceedings to improve the technical quality of the expert reports.

196. The main **conclusions** of this Guide are:

    i. The quantification of harm requires **a dedicated and specific study** of the magnitude of the effects caused by the infringer on the plaintiffs, based on contrasting the actual facts with a reasonable approximation of the counterfactual. **A description must be presented of how the anticompetitive conduct has generated the particular harm** (the theory of harm) that is being quantified. The main objective is to achieve full compensation for any harm suffered as a result of competition law infringements, avoiding both overcompensating and undercompensating the injured parties.

ii. Expert reports submitted in support of claims must be based on solid understanding of the infringement, the sector and the market affected. It must construct the counterfactual scenario based on **transparent, reasonable and technically sound hypotheses, using reliable and verifiable data**.

iii. In addition to having sufficient high-quality data, it is equally important that the data is **processed appropriately and that this is explained in detail.** To increase the transparency of the methodology used and to enable its replicability, it is advisable to include a description of the variables and to provide access to the data, codes, commands and programming procedures utilised, in a processable format, to all parties involved in the judicial process.

iv. **The most commonly used methodologies** in the field of competition law damages claims are:

1. Comparative methods

- **Comparisons of different time periods or diachronic comparisons**, which consist of comparing the evolution of the variable of interest to quantify the harm during the infringement period with **the evolution of the same variable in a period before or after** the anticompetitive conduct.

- **Comparisons of different markets and products or synchronous comparisons**, which consist of comparing the variable of interest during the infringement period with observations of that variable for the: (i) the same product in similar geographic markets not affected by the anticompetitive conduct (**geographical comparison**) or (ii) the same geographical market for similar products that have not been affected by the anticompetitive conduct (**product comparison**).

- **Difference-in-differences method**, which examines the evolution of the variable of interest in the infringing market over a given period, covering the period of the infringement together with the period before or after the infringement, and compares this with the evolution of the same variable over the same period in an unaffected comparator market (i.e., it **combines the diachronic method with the synchronous method**).

2. **The cost-based method and financial analysis**, which consists of calculating a **reasonable and likely value** of the variable of interest that would have resulted had there been no anticompetitive

infringement and comparing this with the value actually observed for that variable of interest, in terms of costs or profitability.

3. **Simulation models**, which, starting from economic theory (industrial organisation models) and including real data (underlying economic relationships), seek to simulate and predict the **functioning of the market**.

v. To reduce the uncertainty inherent in the counterfactual construction and the harm quantification, it is advisable to adopt several **methodological safeguards** when designing and implementing the various harm quantification methods to ensure that the results are **robust and consistent**.

vi. It is advisable for the expert report to explain in great detail how the quantification result has been arrived at, as well as the robustness of the model constructed.

vii. Finally, it should be noted that quantifications based exclusively on harm **estimates** from previous judgements should be limited to those cases where a sufficient degree of similarity is found accompanied by **arguments and evidence** for why that judgement is used as a reference. On the other hand, estimates based on the **automatic application** of an average percentage from **previous** infringements or from **the economic literature may lead to significant errors, without prejudice to their being considered as references**. Each claim, even if it involves the same conduct as another, may have **particularities** that require the quantification method to be adapted to the claim in question.

4. REFERENCES

Angrist J.D. and Pischke J.S. (2008), "Mostly Harmless Econometrics: An Empiricist's Companion", Princeton University.

Baker, J.B. and Rubinfeld, D. (1999), "Empirical Methods in Antitrust Litigation: review and Critique", American Law and Economic Review.

Bolotova, Y. (2009), "Cartel overcharges: An empirical analysis", Journal of Economic Behavior & Organization.

Bönisch, P. and Roman Inderst, R. (2019), "Overcharge Estimation: Making Statistical Evidence More Meaningful", Journal of European Competition Law & Practice, 2019, Vol. 10, No. 8.

Bönisch, P. and Roman Inderst, R. (2021), "Using the Statistical Concept of "Severity" to Assess the Compatibility of Seemingly Contradictory Statistical Evidence (with a Particular Application to Damage Estimation)", Journal of Competition Law & Economics, 18(2), 400–416.

Boswijk, H. P., Bun, M. J., and Schinkel, M. P. (2019), "Cartel dating", Journal of Applied Econometrics, 34(1).

Boyer, M. and Kotchoni R. (2015), "How Much Do Cartel Overcharge?", Review of Industrial Organization.

Buccirossi, P. (2010), "Quantification of Damages in Exclusionary Practice Cases", Contribution to the European Commission Economist workshops on quantification of harm (here)

Bueren, E., Hüschelrath, K. and Veith, T. (2016), "Time is Money – But How Much Money is Time? Interest and Inflation in Competition Law Actions for Damages", Antitrust Law Journal, vol. 81, no. 1.

Castellanos, G. and Solano, D. (2017), "Metaanálisis de la relación entre la orientación al mercado y los resultados de la empresa", (here), Universidad Icesi.

European Commission (2003), "Commission Recommendation of 6 of May concerning the definition of micro, small and medium-sized enterprises" (here)

European Commission (2013), "Communication from the Commission on quantifying harm in actions for damages based on breaches of Article 101 or 102 of the Treaty on the Functioning of the European Union" (here).

European Commission (2015), "Best Practices on the disclosure of information in data rooms in proceedings under Articles 101 and 102 TFEU and under the EU Merger Regulation" (here).

European Commission (2019), "Communication from the Commission – Guidelines for national courts on how to estimate the share of overcharge which was passed on to the indirect purchaser" (here).

European Commission (2020), "Communication from the Commission – Communication on the protection of condifential information by national courts in proceedings for the private enforcement of EU Competition las" (2020/C 242/01).

CNMC (2018), Guía para la presentación de informes económicos en la Dirección de Competencia en la CNMC, (here).

CNMC (2020), Guía sobre el tratamiento de la información confidencial y los datos personales en procedimientos de defensa de la competencia de la Ley 15/2007, (here).

Compass Lexecon (2017), "Economics on demand: Quantifying Competition Damages", (here).

Connor, J. and Bolotova, Y. (2006)," Cartel Overcharges: Survey and Meta-Analysis", International Journal of Industrial Organization.

Connor, J and Lande, R. (2008), "Cartel Overcharges and Optimal Cartel Fines". Issues in Competition Law and Policy, volume 3, ABA Section of Antitrust Law

Connor, J. (2010), "Price Fixing Overcharges: Revised 2nd Edition, SSRN Working Paper (here).

Connor, J. (2014a), "Cartel Overcharges". Research in Law and Economics.

Connor, J. (2014b), "Global Cartels Redux: The Lysine Antitrust Litigation", SSRN, (here).

Cotterill, R. W. and Dhar, T. (2003), "Oligopoly Pricing with Differentiated Products: The Boston Fluid Milk Market Channel", Food Marketing Policy Center Research Reports 074, University of Connecticut.

Daggett, R.S. and Freedman, D.A. (1984), "Econometrics and the law: a case study in the proof of antitrust damages", Technical Report 23, Department of Statistics, University of California, Berkeley, (here).

Davis, P. and Garcés, E. (2009), "Quantitative Techniques for Competition and Antitrust Analysis", Princeton University Press.

Dow, J. (2022), "Pre-Award Interest", in John A. Trenor (Ed.), Damages in International Arbitration Guide (fifth edition), (here).

Droukoupoulos, S., Veronese B. and Witte S. (2020), "Here to stay: regression analysis in follow-on cartel damages", Competition Law Journal.

Finkelstein, M. and Levenbach, H. (1983). "Regression estimates of damages in price fixing cases". Law and Contemporary Problems 46(4):145-169.

Fisher, F. M. (1980). "Multiple regression in legal proceedings". Colum. L. Rev., 80, 702.

Fisher, F. M. (1986), "Statisticians, Econometricians, and Adversary Proceedings". Journal of the American Statistical Association, Vol. 81, No. 394.

Friederiszick, H.W. and Röller, L.H. (2010), "Quantification of harm in damages actions for antitrust infringements: Insights from German cartel cases". ESMT Research Working Papers (here)

Fumagalli, C., Padilla, J. and Polo, M. (2010), "Damages for exclusionary practices: a primer", Competition Law and the Enforcement of Article 102, Chapter 11, Oxford University Press, 2010.

Fumagalli, C., Motta, M. and Calcagno, C. (2018), "Exclusionary Practices: The Economics of Monopolisation and Abuse of Dominance", Cambridge University Press.

García, A., Padilla, J., Watson, N. and Zoido, E. (2018), "Valoración del daño derivado de infracciones de las normas de competencia", in Mariano José Herrador Guardia (Dir.), Responsabilidad civil y seguro. Cuestiones actuales, Lefebvre el Derecho.

Gotanda, J. and Sénéchal, T. (2009), "Interest as Damages", Columbia Journal of Translational Law, Volume 27, Number 3.

Harrington, J. (2004)," Post-cartel pricing during litigation", Working Paper, No. 488, The Johns Hopkins University.

Harris, R.G. and Sullivan, L.A. (1979), "Passing on the monopoly overcharge: a comprehensive policy analysis", U. Pa. L. Rev. 269.

Heller, C. P. and Maier-Rigaud, F. P. (2021), "A practitioners guide to quantifying damage". Economic Analysis in EU Competition Policy. Edward Elgar Publishing.

Hellwig, M. (2006), "Private Damage Claims and the Passing-On Defense in Horizontal Price-Fixing Cases: An Economist's Perspective". J. Basedow (ed.), Private Enforcement of EC Competition Law, Kluwer Law International, Den Haag, 2007.

Hitchings, P. (2010), "Private Enforcement in Spain", Global Competition Litigation Review, 1, 28-40.

Hüschelrath, K., Müller, K. and Veith T. (2013), "Concrete Shoes for Competition – The Effect of the German Cement Cartel on Market Price", ZEW – Centre for European Economic Research Discussion Paper No. 12.

Johnson P., Leamer E. and Leuzinger J. (2017)," Statistical Significance and Statistical Error in Antitrust Analysis", 81 Antitrust L.J. 641.

Kim, D. and Cotterill, R.W. (2008), "Cost Pass-Through in Differentiated Product Markets: The Case of US Processed Cheese", Journal of Industrial Economics.

Kosicki, G. and Cahill, M. (2006)," Economics of cost pass through and damages in indirect purchaser antitrust cases", The Antitrust Bulletin.

Laitenberger, U. and Smuda, F. (2015), "Estimating consumer damages in cartel cases", Journal of Competition Law & Economics.

Lyons, B. (Ed.) (2009). "Cases in European Competition Policy. The Economic Analysis", Cambridge University Press-

Maier-Rigaud, F. and Schwalbe, U. (2013), "Quantification of Antitrust Damages", in David Ashton and David Henry, Competition Damages Actions in the EU. Law & Practice, Elgar, 210-262.

Maier-Rigaud, F., and Sudaric, S. (2019), "The Difference-in-Differences Approach in the Estimation of Cartel Damage". CPI Antitrust Chronicle, 3(1).

McCluer, R. F., and Starr, M. A. (2013). "Using difference in differences to estimate damages in healthcare antitrust: A case study of Marshfield Clinic". International Journal of the Economics of Business, 20(3), 447-469.

McFadden, D., Guthrie S., Liu, P., Wise K. (2003), "The misuse of econometrics in estimating damages", ABA Econometrics Treatise.

Monti, G. (ed.) (2016), "EU law and interest on damages for infringements of competition law: a comparative report", EUI Law Working Paper 2016/11, 271-289.

Motta, M. (2004), "Competition policy: theory and practice", Cambridge University Press.

Niels, G, Jenkins H. and Kavanagh J. (2016), "Economics for Competition Lawyers", Oxford Competition Law, Chapter 9.

Notaro, G. (2013), "Methods for Quantifying Antitrust Damages: The Pasta Cartel in Italy", Journal of Competition Law and Economics.

OCDE (2011), "Quantification of Harm to Competition by National Courts and Competition Agencies". OCDE Competition Committee, Policy Roundtables (here).

OXERA (2003), "Assessing Profitability in Competition Policy Analysis", OFT Economic Discussion Paper 6, July, (here).

OXERA (2006), "Damaged Interest: The Choice of Discount Rate in Claims for Damages", Agenda, September.

OXERA (2009), "Quantifying antitrust damages, towards non-binding guidance for the Courts", study prepared for the European Commission, (here).

OXERA (2015), "Margins of error? Prices vs margins in cartel overcharge estimation",(here).

Prosperetti, L (2009), "Estimating damages to competitors from exclusionary practices in Europe: a review of the main issues in the light of national Courts' experience", Contribution to the European Commission Economist workshops on quantification of harm (here)

RBB Economics and Cuatrecasas (2017), "Study on the Passing-on of Overcharges: Final report", study prepared for the European Commission, (here).

Rubinfeld, D. L. (2008), "Quantitative Methods in Antitrust", Issues in Competition Law and Policy 723.

Rubinfeld, D. L., and Steiner, P. O. (1983). Quantitative methods in antitrust litigation. Law & Contemp. Probs., 46, 69.

Ruiz Peris, J. I., Martorell Zulueta, P., and Marti i Miravalls, J. (2021). Daños y competencia: revisión de cuestiones candentes. Daños y competencia, 275-297.

Seixas, R. and Lucinda, C. (2019), "Computing Cartel Overcharges: when theory meets practice", Estudos Econômicos (São Paulo).

Siotis, G. and Martínez Granado, M. (2010), "Sabotaging Entry: An Estimation of Damages in the Directory Enquiry Service", Review of Law & Economics.

Smuda, F. (2012), "Cartel Overcharges and the Deterrent Effect of EU Competition Law), Centre for European Economic Research (ZEW) Discussion Paper Nº12-050.

Supreme Court (Tribunal Supremo), Sala de lo Civil, "Sentencia 123/2015, 4 de marzo de 2015, ECLI:ES:TS:2015:669" (here).

Tribunal Mercantil de Barcelona, Sección de Derecho de la Competencia (2019), Protocolo de protección del secreto empresarial en los Juzgados Mercantiles, (here).

Vanssay, X. and Erutku, C. (2011), "Damage at the Pump: Does Punishment Fit the Crime?", Journal of Industry, Competition and Trade.

Veljanovski, C. (2019)," Damages for Bid-rigging–The English High Court's Idiosyncratic Cost-Based Approach in BritNed", Journal of European Competition Law & Practice.

Verboven, F. and Van Dijk, T. (2009), "Cartel damages claims and the passing-on defense", Center for Economic Studies, Discussion Paper Series.

Wooldridge, J. M. (2009). "Introducción a la econometría – un enfoque moderno", 4ª edición, Cengage Learning.

Wooldridge, J. M. (2019). "Introductory Econometrics: a modern approach, 7th edition". Cengage Learning.

ANNEX 1: GLOSSARY OF TERMS

Below is a series of terms used in the Guide, in order to facilitate its reading and understanding. Although the terms may have other meanings, the definitions included in this glossary are framed in the context of harm quantification for infringements of competition law.

Actual loss

Decrease in a person's wealth caused by an infringement of competition law.

Adjusted Coefficient of Determination (Adjusted R2)

Indicator of goodness of fit of a linear model whose value, unlike that of R2, only increases with the inclusion of an additional independent variable that adds some explanatory power to the model. Furthermore, with increasing sample size, the values of R2 and Adjusted R2 become closer together.

Arithmetic mean

The average value of a set of numerical data, calculated as the sum of the set of values divided by the total number of values.

Autocorrelation

Autocorrelation measures the relationship between the current value of a variable and its past values. It can occur in time series or panel data models when there is dependence or correlation between errors from different time periods (serial autocorrelation) or geographical areas (spatial autocorrelation).

Autoregression

An estimation method in which the dependent variable depends on its past values.

Barrier to entry

An obstacle or impediment (technological, natural, regulatory, strategic, etc.) that makes it difficult for new companies or operators to enter a market or sector.

Bertrand model

Representation of oligopolistic competition in which companies maximise their profits by choosing their price and taking as given the price of the rest of the competitors.

Bias

The difference between the expected value of an estimator and the population value it is intended to estimate. In the absence of bias, an estimator is centred (on average) on the true population value that is intended to be estimated.

Calibration

The process by which the parameters of a model are adjusted to make them consistent with economic theory or other empirical evidence and then to assess whether its main predictions are consistent with actual observed data.

Capitalisation

The process by which a present or past value is converted into an equivalent future value based on an interest rate. Depending on whether the rate for each period has been calculated only for the initial capital or also taking into account the interest accrued from previous periods, the capitalisation is referred to as simple or compound, respectively.

Coefficient

Parameter which, in the context of an econometric model, represents the average changes that occur in the dependent variable in the face of changes in an independent variable while holding all other variables in the model constant.

Coefficient of variation (Pearson's)

Measure of statistical dispersion calculated as the ratio of the standard deviation to the mean. The higher its value, the more spread out the distributions.

Complementary goods

Those goods that must be used together with other goods to satisfy consumer demand (e.g., a printer and the ink cartridges needed to print). Formally, these are goods with negative cross-price elasticity (if the price of a complementary good increases, demand itself decreases).

Confidence interval

A range of values, derived from sample statistics, between which the true value of a parameter of interest is expected to lie with a specified probability (usually 90%, 95% or 99%).

Consistency

Property of certain estimators by which the bias tends towards zero as the size of the sample increases.

Constant

Parameter of an econometric model that indicates the average value of the explained variable if all explanatory variables were equal to zero.

Control variable

An explanatory variable whose effect on the dependent variable is not the main interest of the analysis, but which is included in the model to take into account its possible influence on the dependent variable. For example, this is often the case of supply and demand variables that are included into a harm quantification regression.

Correlation coefficient

Statistical measure that quantifies the degree of joint variation between two variables. Its value ranges between 1 and -1, being positive when both variables tend to evolve in the same direction and negative when they have opposite dynamics. If the value is 0, their fluctuations are not related, unless there is a third variable that interferes in this relationship.

Correlation matrix

A table showing the correlation coefficients of a set of quantitative variables.

Cost of capital

Cost of financial resources used in a business or investment project.

Counterfactual

The hypothetical situation that would have foreseeably occurred in the absence of the anticompetitive conduct.

Countervailing power of demand

The relative strength of plaintiffs in the process of setting prices and other contractual terms and conditions. This bargaining power may limit the emergence or extent of restrictive competitive practices by bidders.

Cournot model

Representation of oligopolistic competition in which firms maximise their profits by choosing the quantity produced and taking as given the quantity of all other competitors.

Cross-sectional data

Observations of a set of units (e.g., individuals, companies, countries, markets) corresponding to a given moment in time.

Descriptive statistics

Set of metrics that attempt to summarise, order and explain the main characteristics (measures of central tendency, dispersion and position) of a data set.

Determination coefficient (R2)

Indicator of goodness of fit that measures the proportion of variance of the dependent variable that is explained by the independent variables included in the model. Its value ranges from zero to one. An R2 equal to zero would imply that none of the explanatory variables help to explain the variation of the dependent variable, while if it is equal to one it would imply that the explanatory variables perfectly capture said variation. Its value tends to increase the greater the number of variables included in the model.

Discount rate

The cost of capital or interest rate that is applied to determine the present value of a future amount.

Dummy variable

Variable used to incorporate qualitative values into the analysis. It is a binary variable because it can only take the values "1" or "0".

EBITDA

Financial indicator reflecting gross operating profit before financial charges (earnings before interest, taxes, depreciation, and amortisation).

Economic cycle

Fluctuations in economic activity over time, manifested by expansions and contractions in output and other macroeconomic aggregates (employment, investment, general price level, etc.).

Econometric regression

A method that attempts to reflect the impact of changes in one or more explanatory (independent) variables on an explained (dependent) variable, keeping the values of the other explanatory variables constant.

Econometric technique

Combination of economic theory with statistical or quantitative methods to identify and measure relationships between variables.

Economic profitability

Profit or gain associated with an investment.

Efficiency

In econometric terms, this is a characteristic of an estimator referring to the size of its sample variance. Efficiency will be lower the larger the variance, reducing the confidence that the estimate of a parameter obtained from the sample approximates the value of that parameter in the population.

Endogeneity

Existence of a correlation between one explanatory variable and the error term. This phenomenon arises when there are elements included in the error term that are related to explanatory variables included in the model. The existence of endogeneity results in biased and inconsistent Ordinary Least Squares (OLS).

Error term

Information that is not directly explained by the independent variables and incorporates randomness into the model.

Estimation

A set of statistical and econometric techniques that attempt to approximate the population value of a parameter from a sample.

Expected value

The average value of a random variable.

Explanatory, dependent or endogenous variable

The variable that is explained by using a regression model.

Explanatory, independent or exogenous variable

The variable by which the behaviour of the dependent variable is intended to be explained.

Extrapolation

A procedure by which the value of a variable is estimated beyond the available data range, according to its relationship with other variables.

F-statistic

A statistic commonly used in multiple regression models to assess the joint explanatory power (significance) of a group of independent variables on the explained variable.

First differences

A transformation carried out on a database with a time dimension consisting of subtracting from each variable its value in the immediately preceding period.

Fixed cost

The cost that does not vary depending on the quantity produced.

Fixed effects model

A model with panel data in which it is assumed that unobserved heterogeneity may be correlated with an explanatory variable. Normally, fixed effects related to time, geographical scope or the market in question are included.

Follow-on action

Process of claiming damages based on a competition authority finding an infringement of Articles 1 or 2 of the LDC and/or 101 or 102 of the TFEU.

Generalised Least Squares (GLS)

Econometric estimation method that takes into account the existence of heteroscedasticity or error autocorrelation with a known structure, giving greater weight to those observations that present less variance in the error term.

Goodness of fit

The degree of closeness between the values predicted by a model and the observed values. An example of a goodness-of-fit measure is the coefficient of determination (R2).

Goodwill

A value, which refers to a company's capacity to generate profits thanks to intangible assets such as brand value, market positioning or customer base.

Gross profit

The difference between total sales and the costs directly related to those sales for a specific point in time, before the application of taxes, depreciation and other deductions. It is a measure of the ability to obtain results directly linked to a given activity.

Hausmann's test (or Durbin-Wu-Hausman)

A contrast used to determine whether the differences between two estimates are significant. It can be used for various purposes such as evaluating the consistency of an estimator or the relevance of a variable.

Heteroscedasticity

A situation that occurs when the error term has a variance that is not constant across observations and over time. Heteroscedasticity leads to OLS estimators becoming inefficient although they are still unbiased and consistent.

Homoscedasticity

This characteristic of a regression model is true if the error term has a variance that is constant over time and independent of the value of the explanatory variables. When these conditions are not met, it is termed heteroscedasticity.

Hypothesis testing

Statistical procedure aimed at assessing whether certain assumptions about the parameters estimated for a population are compatible with the information contained in the sample.

Income elasticity of demand

Variation in the quantity demanded of a good or service in response to changes in consumer income (without changing prices).

Instrumental variable

A variable, which is usually used to solve endogeneity problems. When a model has an endogenous explanatory variable, the instrumental variable does not appear in the model, it is independent of error and correlates with the endogenous variable.

Interest in arrears (Moratory interest)

The amount of compensation payable to the debtor of an obligation for the delay in fulfilling it.

Internal rate of return (IRR)

The rate of interest that would make the net present value (NPV) of a project or the company being valued zero. This method is frequently used to measure profitability.

Interpolation

Approximation of the value of a variable from earlier and later data. In its simplest version (linear interpolation), a line is drawn between two points.

Lagged variable

The variable, which refers to past periods that influence the explained variable at present.

Legal interest rate

The percentage rate, used to calculate interest, which is not determined by an agreement between the creditor and the debtor but by legal provision, usually the General State Budget Law.

Loss of profit

An increase in the plaintiff's wealth that would have occurred in the absence of the competition law infringement.

Marginal cost

Additional cost incurred by increasing production by one unit.

Median

Central value of a data set of one variable, that is, the value that divides the set into two equal parts.

Mode

The value that appears most frequently in a data set.

Monopolistic competition model

A market represented by a high number of companies with differentiated products, but close substitutes, and low barriers to entry. Differentiation provides each firm with some market power, which allows it to raise the price above the marginal cost.

Moving average

Arithmetic average of a certain number of data points prior to each period (the set of data on which the average is made varies over time, hence it "moves"). It is often used to smooth out fluctuations in the data, with the result varying depending on factors such as the number of periods used for calculation it or whether weights are included.

Multicollinearity (imperfect)

A term indicating that the correlation between some explanatory variables in a model is high. As a consequence, OLS estimators will no longer be efficient and it may be difficult to estimate the individual effect of the affected variables. Nevertheless, the estimators will remain unbiased and consistent.

Net Present Value (NPV)

A valuation method that consists of discounting the value of the future cash flows of a company or investment project, using an appropriate discount rate.

Net profit

The result of deducting other expenses (mainly financial and tax expenses) from operating profit.

Nominal and real variable

A nominal variable is a variable expressed in nominal monetary terms (e.g., current euros), while a real variable has a monetary value expressed with respect to a base period (e.g., constant euros). To convert a nominal variable into a real variable it is necessary to divide it by a price index.

Normal distribution

A Bell-shaped (or Gaussian) probability distribution that is symmetrical with respect to its mean; it is often used in statistics and econometrics to model a population.

Null and alternative hypothesis

In hypothesis testing, the null hypothesis (H0) is a condition that is taken to be true and which assumes that the parameter takes a certain value. An alternative hypothesis (H1), which is the opposite proposition, assumes that the parameter has a value different from the one assumed in the null hypothesis (H0).

Omitted variable

A variable that does not appear in the model as an explanatory variable yet influences the dependent variable or other explanatory variables (that is why it is in the error).

Operating profit

The result of subtracting operating expenses from gross profit.

Ordinary Least Squares (OLS)

Econometric method for estimating the parameters of a linear regression model. Estimates are obtained by minimising the sum of the squared residuals.

Outliers

Observations in a database that are substantially different from the distribution of the rest of the data. This can be due to several reasons, including, for example, errors when creating the database or data from a different population.

Overcharge pass-on

A situation that occurs when an agent that has suffered harm (competitor, supplier or purchaser) caused by an infringement of competition law passes on part or all of the harm suffered to its direct purchasers, reducing or even eliminating that harm.

P-value

The minimum significance level at which the null hypothesis can be rejected (e.g., that a behaviour has had no effect). As it is a probability, its value is between 0 and 1.

Panel data

Data structure combining information from several individuals at several points in time (e.g., price data over ten years from five companies within a cartelised market). The main feature of this data structure is that the units observed over time are always the same (e.g., the same companies before, during and after the infringement).

Parameter

A numerical value that describes certain characteristics of a population. It is normally an unknown value that is estimated using statistical inference techniques.

Perfect competition model

It represents a market characterised by a large number of sellers and buyers, a homogeneous product, no barriers to entry and exit, perfect information, and agents with no ability to individually influence the market price, which will be equal to marginal cost.

Population

The set of all similar elements that are of interest to a study or estimation.

Price effect

This occurs when anticompetitive conduct causes buyers to pay higher prices for each unit of the affected product purchased than would otherwise be the case. In the case of a purchasing cartel, the effect would consist of "under-invoicing" by the suppliers for each unit sold of the affected product.

Price elasticity of demand

A variation in the quantity demanded of a good or service due to changes in the price of that same good or service (own price elasticity) or another (cross-price elasticity).

Price index

Statistical measure that shows the evolution over time of the prices of certain goods and services. One of the most widely used is the CPI (Consumer Price Index).

Principle of effectiveness

Principle enshrined in Directive 2014/104/EU (Article 4) whereby national requirements on the quantification of harm in cases of competition law infringements must not make it impossible or excessively difficult to exercise the EU right to compensation for harm.

Principle of equivalence

Principle enshrined in Directive 2014/104/EU (Article 4) whereby national requirements on the quantification of harm in cases of competition law infringements must not be less favourable than those governing similar national actions.

Principle of indemnity

Principle requiring full compensation for the harm caused, whereby economic compensation must be aimed at restoring the situation to what it was at the time of the harm, whereby compensation must be adjusted as far as possible to the purchasing power of the amount to be received.

Profit margin

The difference between the selling price of a product and the cost of producing or purchasing it.

Proxy

The variable that is related to but not identical to another variable of interest, which is why it is usually used as an approximation if the latter is not available.

Random effects model

A model with panel data in which it is assumed that unobserved heterogeneity is not correlated with explanatory variables.

Range

A numerical value indicating the difference between the maximum and minimum value of a population or statistical sample. As it is highly dependent on outliers or extremes, the interquartile range, which is the difference between the third (Q3) and the first quartile (Q1), is generally used as a measure of dispersion.

Replicability

A common concept in the scientific method that refers to the potential for a study to be reproduced by another expert in order to check the validity of the calculations and results.

Residue

The observed difference between the actual value of the explained variable and the value predicted by an econometric model for each observation in the sample.

Risk-free interest rate

The return that would be obtained by investing in an asset for which the level of risk is virtually non-existent (e.g., government bonds). It is sometimes used as the minimum threshold required for an investment.

Robustness

A characteristic of the results that occurs when their validity is not affected by small changes in the starting assumptions, which can be made in the framework of a sensitivity analysis.

Sample

A selected subset of data belonging to a population.

Seasonality

Periodic and predictable dynamics of certain variables that are repeated every particular period of time, normally equal to or less than a year. This has to be taken into account when estimating whether the frequency of the data used is greater than annual.

Sensitivity analysis

The process of assessing how changes in a model (inclusion, exclusion or transformation of variables, modification of the time period, elimination of potential outliers, etc.) influence its results.

Significance level or significance

Probability of committing a type I error (false positive) when testing a hypothesis.

Simultaneity

A term that implies that one or more explanatory variables of a regression model are determined together with the dependent variable (e.g., the price and quantity of a certain product).

Spurious correlation

Existence of a high correlation between two variables without a causal relationship between them.

Stand-alone action

Damages claim process that is not based on a competition authority finding an infringement of Articles 1 or 2 of the LDC and/or 101 or 102 of the TFEU.

Standard error of an estimator

Value showing the dispersion of the sampling distribution of an estimator (e.g., the parameters of a regression model). It is used to measure the accuracy of the estimate. In general, the greater the standard error, the less precise the estimate.

Standard or typical deviation

A measure that provides information on the dispersion of a variable, usually with respect to its mean. It is obtained by taking the square root of the variance and is always positive.

Statistical independence

Two variables are statistically independent when the movements of one do not affect those of the other.

Statistical inference

Set of techniques used to approximate the behaviour of a population based on information provided by a sample of that population.

Statistical significance

The probability that the outcome of an estimate is not due to chance. It is therefore a criterion which, based on the hypothesis tests, allows statements to be made about the estimated values of the parameters (β_1, β_2, …).

Statistical technique

Mathematical methods for the collection, analysis and interpretation of a data set.

Structural change

Structural change occurs when the value of one of the parameters in a regression model changes suddenly over time (e.g., when there is a technological breakthrough or if there is a severe economic downturn). Possible tests for structural change include the parametric Chow test or the CUSUM (cumulative sum) test.

Substitutable goods

Goods that can satisfy the same need as others and which are therefore considered to be substitutable (e.g., sugar and sweetener). Formally, they are those with positive cross-price elasticity (if the price of a substitute good increases, demand increases).

Sunk cost

All those costs that have already been incurred and cannot be recovered.

T-statistic

A statistic commonly used in regression models to assess the individual explanatory power (significance) of an independent variable on the explained variable.

Time series data

Observations of a single variable (e.g., GDP, price index, etc.) at certain times (days, weeks, months, years, etc.), which are ordered chronologically.

Time value of money

Evolution of the value of monetary flows due to the passage of time. Usually, monetary amounts lose value over time due to various factors such as inflation or unrealised investment opportunities (opportunity cost). For this reason, money flows at different points in time are not directly comparable unless updating or capitalisation operations are carried out.

Total cost

The sum of fixed costs and variable costs.

Trend

Long-term movement of a time series. This can be approximated by including an explanatory variable reflecting the time dimension.

Two-Stage Least Squares (2SLS)

An econometric estimation method that is often used to correct for an endogeneity problem in a model by applying instrumental variables. In these cases, the method allows consistent estimates to be obtained in comparison with OLS, provided that relevant instrumental variables (correlated with the endogenous explanatory variable) and exogenous (not correlated with the error term) variables are used. In return, they tend to be less efficient estimates (standard errors are usually larger).

Type I and Type II errors

When testing a statistical hypothesis, two types of errors can be made:

Type I error or false positive: the null hypothesis is rejected when it is, in fact, true at the population level. For example, the null hypothesis that a behaviour has had no effect is rejected, and therefore it is concluded that there has been an effect, when in fact there has been no effect (hence the "false positive").

Type II or false negative error: the null hypothesis is not rejected even though it is false. In the previous example, we fail to reject the null hypothesis that a behaviour has had no effect, and therefore conclude that there has been no effect, when in fact there has been an effect (hence the "false negative").

Umbrella effect

A phenomenon whereby companies that do not engage in anticompetitive conduct, but which sell substitute products, consciously or unconsciously charge higher prices by taking advantage of the existence of the infringement.

Unobservable heterogeneity

In a panel data model, this concept refers to that part of the error term that does not vary over time among the individuals or groups considered. Depending on the assumptions made about their relationship with the rest of the explanatory variables, fixed-effects or random-effects, models can be used to control their impact on the estimates.

Updating or discounting

A process by which a future value is converted into an equivalent present value based on a discount rate.

Variable cost

The cost that varies depending on the quantity produced. This is often used as an approximation of marginal costs.

Variance

A dispersion measure of the distribution of a random variable. Its value is always positive and corresponds to the arithmetic mean of the squares of the deviations from the mean (i.e., it is equal to the standard deviation squared).

Variance Inflation Factor (VIF)

A measure that quantifies the strength of multicollinearity in a regression analysis by OLS. It provides an index that measures the extent to which the variance of an estimated regression coefficient is increased due to correlation with other explanatory variables.

Volume effect

This occurs when a purchaser of the product affected by anticompetitive conduct passes on part of the overcharge to their purchasers, giving rise to decreased sales, which may translate into lower profits compared to the situation with no infringement.

Weighted average

A measure of central tendency, which is obtained from a data set with different levels of importance for the analysis to be carried out. To calculate this, each datum is multiplied by its importance (or weight), added together (this is called a weighted sum) and, finally, the figure obtained is divided by the sum of the weights.

Weighted Average Cost of Capital (WACC)

Average of the cost of the two capital resources that a firm has (debt and equity), weighted by their relative weights in total liabilities. While the cost of debt is usually easier to obtain (considering the interest paid to creditors), the cost of equity must be estimated using several methods, including the Capital Asset Pricing Model (CAPM) or the Arbitrage Pricing Theory (APT).

Weighted Least Squares

A particular case of GLS, which is used to adjust for heteroscedasticity, weighting the observations by the inverse of the variance of the error (greater weight is given to those observations that have less variance in the error term).

ANNEX 2: STATISTICAL AND ECONOMETRIC CONCEPTS

197. This annex aims to collect and facilitate the understanding of several relevant statistical and econometric concepts relevant in quantifying harm. This is not an exhaustive review of all the concepts, for which specialised manuals are recommended [107].

A.2.1. STATISTICAL CONCEPTS

A2.1.1 Data types

198. The common denominator of all the analyses and techniques introduced in this Guide is their application to a dataset that contains relevant information under a specific structure, often marked by the availability of the data itself.

199. The structure of the available data is relevant to the extent that it conditions the type of analysis that can be performed. The following data structures are highlighted below:

- **Cross-sectional data**: These are observations of several individuals (e.g., consumers, users, companies) or variables (e.g., prices, margins, costs) at a given point in time (e.g., in a particular year, month, week, day).

- **Time series data:** These data contain observations of a single or individual variable (e.g., GDP, price index, etc.) over time (days, weeks, months, years). Time series data, in comparison to cross-sectional data, include the time dimension (converting them to dynamic data), which allows for the consideration of potential influence of past events on future ones, as well as possible "lagged effects" in the impact of certain behaviours. However, they are often more challenging to analyse because of the frequent dependence of variables over time, the existence of trends, or seasonality[108].

- **Panel data:** This combines elements of the two previous structures as it contains information from several individuals over time (e.g., price data over several years for all companies belonging to a market that was cartelised). The main feature of this structure is that the units observed over time are always the same (e.g., the same companies before, during

[107] See, for example, Wooldridge (2019) and Angrist and Pischke (2008). For an approach more adapted to competition issues, consult Chapter 2 of Davis and Garcés (2009).

[108] For more information, see Section 6 of this Annex.

and after the infringement)[109]. Although the availability of this type of data can be complex, it has advantages over other data structures that contain a time dimension because it allows for controlling unobservable heterogeneity[110] given the information on other units over time.

- **Cross-section merged data (pooled data):** This structure builds on the characteristics of panel data, with the difference that the selection of units at each time point (e.g., monthly average price) is random. Therefore, the observations (e.g., prices), although they always belong to the same set (e.g., a certain geographical market that was cartelised), are not necessarily the same for each moment (e.g., month).

A2.1.2 Statistical parameters

200. Throughout this section, several parameters that may be useful for carrying out a harm quantification exercise are described, as are some of the most common ways of representing them graphically.

201. To do this, a simple example based on dummy data will be used. Let's assume we have data on the prices of a product applied by 32 firms in two markets at a given point in time. We will also assume that one of the two markets is cartelised ($Price_c$) and the other is not ($Price_{nc}$). Below is a table with price data for both markets that will be used to calculate the main descriptive statistics detailed in the next section, as well as the summary of these statistics.

[109] Panel data can be "balanced" when you have observations for all individuals throughout all time periods included in the research, or they may be unbalanced when there are periods without data for some individuals in the study.

[110] See Section A2.6.3. for more information on this.

Table 2. Prices by market (left) and summary statistics for the baseline scenario (right)

Firm	$Price_c$	$Price_{nc}$	Cartel Costs
1	9,0	8,5	6,6
2	12,2	9,0	6,1
3	13,1	10,2	10,9
4	14,1	11,4	9,6
5	14,2	11,5	7,1
6	14,4	11,7	7,2
7	15,1	11,8	12,1
8	15,6	11,9	13,9
9	16,5	12,5	14,9
10	17,0	12,6	13,6
11	17,6	12,8	14,1
12	17,8	13,2	13,0
13	17,9	13,5	13,1
14	17,9	13,5	13,1
15	18,2	13,5	14,5
16	18,5	13,5	14,8
17	18,7	13,6	9,3
18	19,0	13,7	15,2
19	19,3	16,0	14,1
20	19,5	16,1	15,6
21	19,8	16,3	14,4
22	19,8	16,5	14,4
23	19,8	17,2	15,8
24	20,5	17,2	16,4
25	20,7	17,5	8,3
26	20,8	17,9	15,2
27	21,2	18,2	20,1
28	22,1	18,5	11,1
29	23,2	19,6	20,9
30	24,0	19,7	21,6
31	26,0	20,7	15,6
32	28,0	21,1	18,5

Statistic	$Price_c$	$Price_{nc}$
Mean	18,5 €	14,7 €
Median	18,6 €	13,6 €
Mode	19,8 €	13,5 €
Variance	15,6	11,4
Standard deviation	3,9 €	3,4 €
Coef. Of variation	0,21	0,23
Q1	15,8 €	12,1 €
Q2	18,6 €	13,6 €
Q3	20,7 €	17,4 €
Maximum	28,0 €	21,1 €
Minimum	9,0 €	8,5 €
Range	19,0 €	12,6 €
Interquartile Range	4,8 €	5,4 €

Source: prepared in-house.

A2.1.2.1 Descriptive Statistics

202. Descriptive statistics allow us to synthesise the information contained in the data sets. As such, the following statistics can be distinguished by categories.

Measures of central tendency

203. The **arithmetic mean**[111] is the sum of a set of values divided by the total number of values. The arithmetic mean is the average value of the set of data being analysed. While the arithmetic mean is the most commonly used statistic, as it best represents the data if it is normally distributed, it is important to note that it is very sensitive to outliers or extremes, as will be shown later in Subsection A2.1.3.2.

$$Mean = \frac{Price\_C_{E1} + Price\_C_{E2} + \cdots + Price\_C_{E32}}{32\ companies} = 18.5\ €$$

[111] In addition to the simple arithmetic mean, there is another type of mean such as the weighted average.

204. **The median** is the 'central' value of a variable. To calculate it, it is necessary to arrange the data series observations (n) in increasing or decreasing order, with the median being the value (X_n) that divides the series into two equal parts. If the number of data is even, as in our example, the median is the average of the two values in the middle of the series.

$$Median\ if\ n\ odd = X_{\frac{n+1}{2}}$$

$$Median\ if\ n\ even = \frac{1}{2}\left(X_{\frac{n}{2}} + X_{\frac{n}{2}+1}\right)$$

205. Applying the formula to our example for cartelised prices, with X_n being the number of companies in the series, ordered from smallest to largest, and n being the number of observations, the median is calculated:

$$Median = \frac{1}{2}\left(X_{\frac{32}{2}} + X_{\frac{32}{2}+1}\right) = \frac{1}{2}(X_{16} + X_{17}) = \frac{1}{2}(18.5 + 18.7) = 18.6€$$

206. A simple and preliminary way to analyse the distribution in a data set is to compare the **mean and median**. The greater their difference, the more likely it is that we are dealing with an **asymmetric** data series, in which there could be outliers. In the case analysed, we see that the differences are small in both markets, being smaller in the cartelised market (0.1 euros) than in the non-cartelised market (1.1 euros).

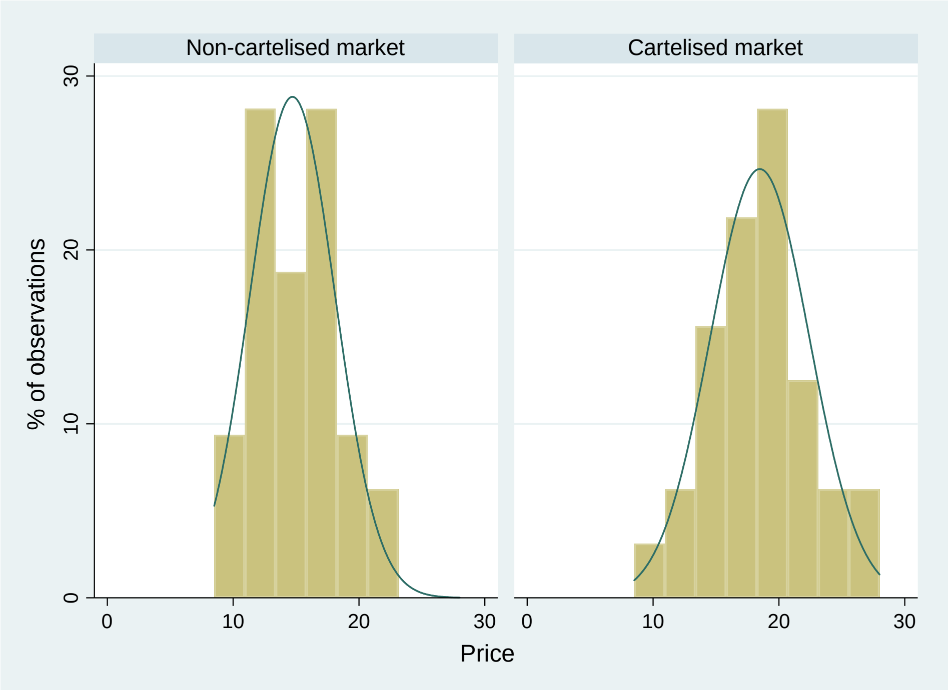
207. The **mode** is the most frequently repeated value in the distribution. In our example, the mode is 19.8 euros for the cartelised market and 13.5 euros for the non-cartelised market. If the distribution were perfectly symmetrical (e.g., normal), the mean, median and mode would coincide.

208. The distribution of a variable can be graphically represented by a **histogram or bar chart**, as shown in Figure 4, in which each bar is proportional to the size of its frequency (absolute or relative) in the distribution. Histograms allow us to approximate the shape of the distribution and compare it with the normal (symmetrical) distribution, which is part of most theoretical assumptions in statistical and econometric analysis.

209. Continuing with our example, the distribution of cartelised prices (relative frequency) more closely resembles a normal distribution (represented by the curve in the graph), while that of non-cartelised prices is more skewed to the right[112].

[112] This is also reflected in the fact that its median is less than its mean.

Figure 4. Price histograms for the non-cartelised and cartelised markets



Source: prepared in-house.

Measures of non-central tendency

210. These measures divide the data series into equal parts and serve to rank an individual within a given sample or population. They require the observations to be arranged in increasing or decreasing order.

211. **Quartiles** are measures of location that divide the population into four equal parts. The first quartile (Q1) is the value that has 75% of the values above it; the second quartile (Q2) has 50% of the values above it and coincides with the median; the third quartile (Q3) has 25% of the variables above it, and the fourth quartile (Q4) coincides with the maximum value of the data series. In other words, given a sample of 100 data points ordered from lowest to highest, the first quartile would be the 25th value in the series, the second quartile the 50th value in the series, the third quartile the 75th, and the last quartile the 100th.

212. In the case of **deciles**, the population is divided into ten equal parts, with the first decile having 90% of the values above it. Continuing with the previous example, assuming we had 100 data points sorted from lowest to highest, the first decile would be the 10th value in the series.

213. **Percentiles** follow the same reasoning. Thus, the first percentile has 99% of the values above it and, hence, the 25th, 50th, and 75th percentiles coincide respectively with the first, second, and third quartiles.

Measures of dispersion

214. In statistical analysis, it is also important to know whether the distribution of the data is close to or far from the central values to determine if they are representative.

215. The **range** is the difference between the largest and the smallest value of a variable. The formula and its application to the cartelised prices in the example are shown.

$$Range = Max_{PC} - Min_{PC} = 28 - 9 = 19\text{ €}$$

216. As it is highly dependent on outliers or extreme values, the **interquartile range**, which is the difference between the third (Q3) and the first quartile (Q1), is generally used as a measure of dispersion. Continuing with cartelised prices, its formula would be as follows:

$$Interquartile\ range_{PC} = Q_3 - Q_1 = 20.7 - 15.8 = 4.8\text{ €}$$

217. The **deviation** is a measure of dispersion that shows the separation between any value in the series and another value in the series, usually the mean.

218. The **variance** is the arithmetic mean of the squares of the deviations from the mean. The square root of the variance, known as the **standard deviation** from the mean[113], is often calculated and has the advantage of being expressed in the same units as the data from which it is calculated (in this example, euros). Both measures, variance and standard deviation, are always positive and indicate the dispersion degree of the analysed data.

219. **Pearson's coefficient of variation** is the ratio of the standard deviation to the mean. Therefore, it will have a higher value the more dispersed the distribution is. Its main advantage is that it enables a comparison of the data series dispersion with different units of measurement.

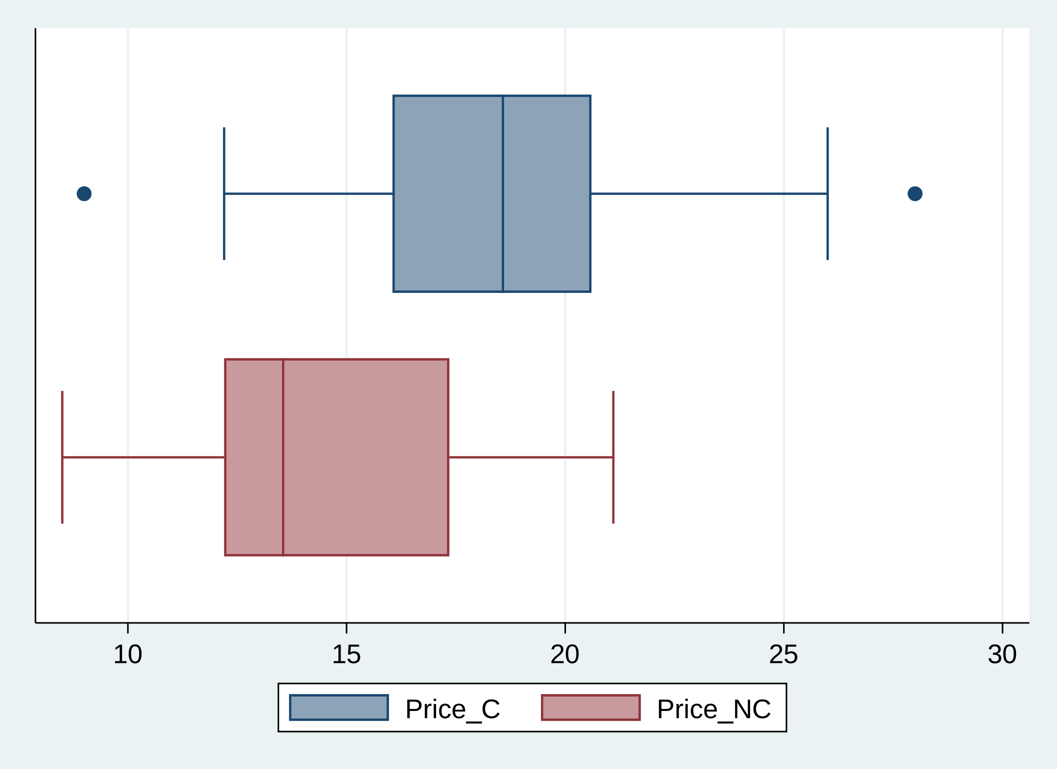
---

113 If the distribution of the data of a variable approaches a normal distribution, which is the most frequently used, it is verified that:

- 68% of its values are located at a distance from the mean of less than one standard deviation.
- 95%, at a distance from the mean of less than two standard deviations.
- 99%, at a distance from the mean of less than three standard deviations.

220. Several of these concepts can be represented by a **box-and-whisker plot**. In this way, a box is presented whose length or height (depending on the orientation of the graph) is given by the interquartile range (which is 50% of the central observations), with a line inside it reflecting the median. It is also common to see a dot or cross inside the box indicating the mean. Figure 5 shows the box-and-whisker plot using the data from the example.

Figure 5. Box and whisker plot for the price of both markets



Source: prepared in-house.

221. Whiskers emerge from each end of the box; their length corresponds to the first and third quartile values, multiplied, if following the traditional Tukey rule[114], by 1.5 times the value of the interquartile range[115]. Values outside

114 See Subsection A2.1.3.2.

115 To clarify the explanation that applies to the two figures, we will take as a reference the price charts (blue box and whiskers in the upper graph).

- To construct the left whisker, we start from the value of Q1 (15.8) and subtract 1.5 times the interquartile range (1.5 x 4.8 = 7.2). Through this subtraction, we obtain a value of 8.6, which is the minimum price up to which the left whisker could reach (the real length is marked by the first price greater than 8.6, in this case 12.1).
- The same operation is conducted to build the right whisker, although in this case starting from Q3 (20.7) and multiplying the RI (7.2) by 1.5. In this case, the two quantities would have to be added together, giving rise to a maximum theoretical value of the right whisker of 27.9 (the real length is marked by the first price below that value, in this case 26).

the diagram are represented by a dot suggesting the possible existence of an outlier[116].

222. If we look at the example of cartelised prices, we can see that the left-hand side of the box is smaller than the right-hand side, indicating that prices between 25% and 50% of the population are more concentrated (less dispersed) than those between 50% and 75%. The same analysis can be performed by comparing the length of the whiskers: the longer the length, the greater the dispersion of the values. The left whisker is shorter than the right one, which means that the lowest 25% of prices are more concentrated than the highest 25%. Finally, a possible outlier is observed on each side, contrary to that seen in the non-cartelised prices.

A2.1.2.2 Correlation between variables

223. In the field of harm quantification, it is particularly important to analyse the relationship between various variables. Notably, it is of interest to know how the conduct of the defendant companies affects the economic performance of the plaintiffs. It may also be important to know whether a change in supply or demand conditions in the market is usually accompanied by changes in the prices or profitability of the companies. These types of questions are usually analysed using the correlation coefficient, which measures how close the relationship between two variables is to a perfect linear relationship[117].

224. The **correlation coefficient** has a value between -1 and 1. A negative value implies that the two variables vary in opposite directions (when one variable increases, the other decreases). A positive value implies a variation in the same direction (both tend to increase or decrease at the same time). If the value is zero, their fluctuations are unrelated (unless a third variable interferes with this relationship).

225. This indicator is commonly used to check how variables are related to one another, together with a visual analysis of scatter diagrams, and can be a preliminary step in selecting which variables to include in a model. However, we must bear in mind the fact that **observing two variables that are strongly correlated does not necessarily imply that there is a causal**

[116] The concept is developed in Subsection A2.1.3.2. In the example of cartelised prices, there are two points outside the diagram, corresponding to 9 and 28 euros, i.e., the extreme values that are outside the range when applying Tukey's rule. On the other hand, in the case of non-cartelised prices, the values 21.1 and 8.5 are not represented with points outside the diagram as they are contemplated within the limits described in the previous footnote and, therefore, are not considered atypical values.

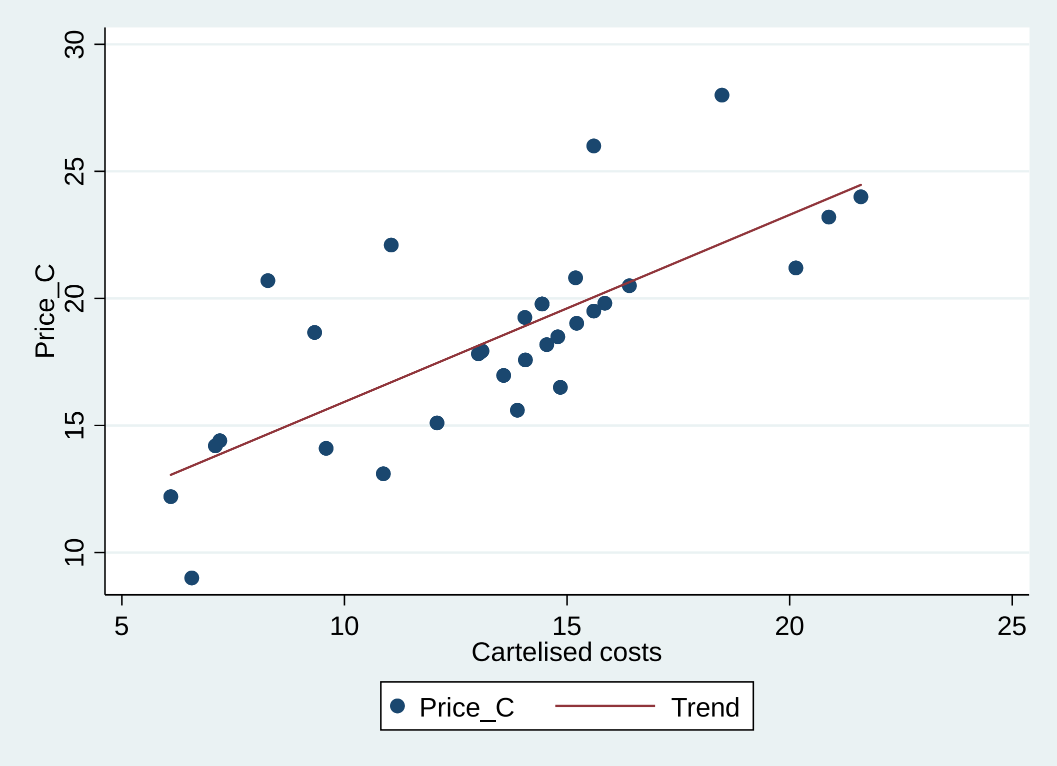
[117] A linear relationship between two variables implies that they both move in the same direction at a constant rate, so the relationship between them can be represented graphically by a line.

relationship between them. For example, two variables may move together purely by chance or because there are in fact other variables that are not being considered that are causing the relationship. This is a fundamental principle of statistical and econometric analysis that should be considered when quantifying harm.

226. The most common way of graphically analysing the type of relationship between two variables is by using a scatter diagram or point cloud, as shown in Figure 6. Each variable is represented on an axis, so it is possible to observe whether they show any kind of co-evolution and, if so, what form it takes. Sometimes straight lines (as in the example) or curves are represented to try to see how they fit the dynamics displayed by the data.

227. In our example, we will assume that we have production cost data for each of the 32 companies. We will plot the cost on the horizontal axis and the cartelised price on the vertical axis.

Figure 6. Scatterplot of Cartel Costs and Prices



Source: prepared in-house.

228. As depicted in the graph, a positive relationship exists between the two variables: when one increases, the other also tends to rise. The correlation coefficient is positive and relatively high (0.77), suggesting the existence of a linear relationship between the two; this seems consistent with what would be expected according to economic theory (if costs increase, prices will

increase). However, on other occasions, the direction of the relationship and its form are less obvious. Moreover, as we will see in Section A2.1.3.2, this measure is very sensitive to the presence of outliers or extreme values.

A2.1.3 Sample analysis

229. When trying to quantify the harm, it is relatively common not to have all the data available for scenarios with and without infringement. In these cases, there is a sample available, which is a subset of data from a reference population. When analysing data from a sample, it needs to be representative, i.e., its characteristics should be close to those of the population under study. Otherwise, the conclusions drawn from the sample analysis will be biased (systematic error).

230. The representativeness of a sample depends on several factors, including the data selection method[118], the transformations carried out (e.g., aggregation or elimination of certain data may compromise its representativeness) or the sample size[119] (in principle, the larger the sample size, the better the representativeness).

231. Below are some problems that may occur relatively frequently in data samples and that may compromise their representativeness, as well as possible approaches (which should always be transparent and reasoned).

A2.1.3.1 Missing values

232. When constructing a database, it is possible that not all observations of the variables used are available. This existence of **missing values** can compromise the representativeness of a sample. The key in these cases is whether the missing information is randomly distributed in the sample or, on the contrary, it mainly affects one category of observations[120] and may lead to bias in the analysis.

233. In this situation, one possibility is to **exclude** all observations with missing information, thereby reducing the sample size; this may affect the results of

---

[118] Statistical and econometric techniques normally start from the assumption that the samples are random. However, the data used for the quantification of harm is usually not the result of a random sample from a larger population but are rather constructed from all the information to which the parties of the procedure have access.

[119] It is important to point out that there is no minimum sample size to be able to carry out a statistical and/or econometric analysis with a certain level of confidence. However, relevant issues such as the precision of the estimates, the significance tests or the confidence intervals may vary depending on the size of the sample.

[120] For example, if the lack of information occurs in all the data in a time period, or in certain brands or models.

the analysis to a greater or lesser extent depending on how they are distributed. Another option is to resort to an **imputation technique**, which consists of replacing missing data with other values. In this area, there is a wide variety of techniques with varying degrees of complexity, for example:

a. One possibility would be to impute missing data using the mean, median, mode or a random value from the sample.

b. In time series, methods such as moving averages, interpolations or linear extrapolations[121] can be used, as well as other more complex methods[122].

234. In general, there is no one technique that is always preferable, it will always depend on the circumstances of each case (importance of the imputed variable, percentage of missing data, etc.). Nevertheless, the rationale for choosing a specific technique should be explained in detail.

235. Finally, a good practice is to show the estimates **with and without missing values** after the use of the above techniques and discussing possible differences in the results.

A2.1.3.2 Outliers

236. Especially when sample sizes are small, it can occur that the results of the analysis performed are very sensitive to the presence of certain observations, usually referred to as outliers or extremes. Sometimes outliers arise as a result of coding errors in the construction of the database, and their value is abnormally lower or higher than the rest. In such cases, the most advisable solution would be to remove them, with due transparency. However, it is not always evident whether a particular observation is an outlier or not, as this is a somewhat subjective concept.

237. Various methods exist for detecting outliers or extremes: employ graphical analysis[123], normalise the variable of interest and identify observations as outliers if they deviate by more than a specified threshold of standard deviation from the mean, or use statistics such as Tukey's test[124] or Cook's

121 While linear interpolation consists of using the data immediately before and after the one to be imputed and joining these with a line, linear extrapolation draws a line from the preceding or subsequent data. The Practical Guide by the European Commission (2013) elaborates both methods in greater detail, using graphics.

122 Such as multiple imputation methods or ARIMA models, among others.

123 Above all, using box and whisker plots.

124 Consider "slight" outliers to be those at a distance greater than 1.5 times the interquartile range from the first and third quartiles (i.e., values outside the "whiskers" of the diagram above). Values that are more than three times that range apart are called “extreme” outliers.

distance[125]. In cases, in which the origin of the outliers is not clear and may be due to the very nature of the data, it is advisable to present the results with and without outliers. This way, the sensitivity to such observations can be analysed, always reasoning the final decision to include or exclude such observations in the proposed estimates.

238. Continuing with our **example**, let's assume that, due to a data entry error, the cartelised price of company 20 (see Table 2) would increase from 19.5 to 195. This leads to important changes in the sample and in the key statistics describing the sample, as can be seen below:

Table 3. Comparison of the main statistics on the cartelised price variable (Price_C) after including an outlier (Price_C*).

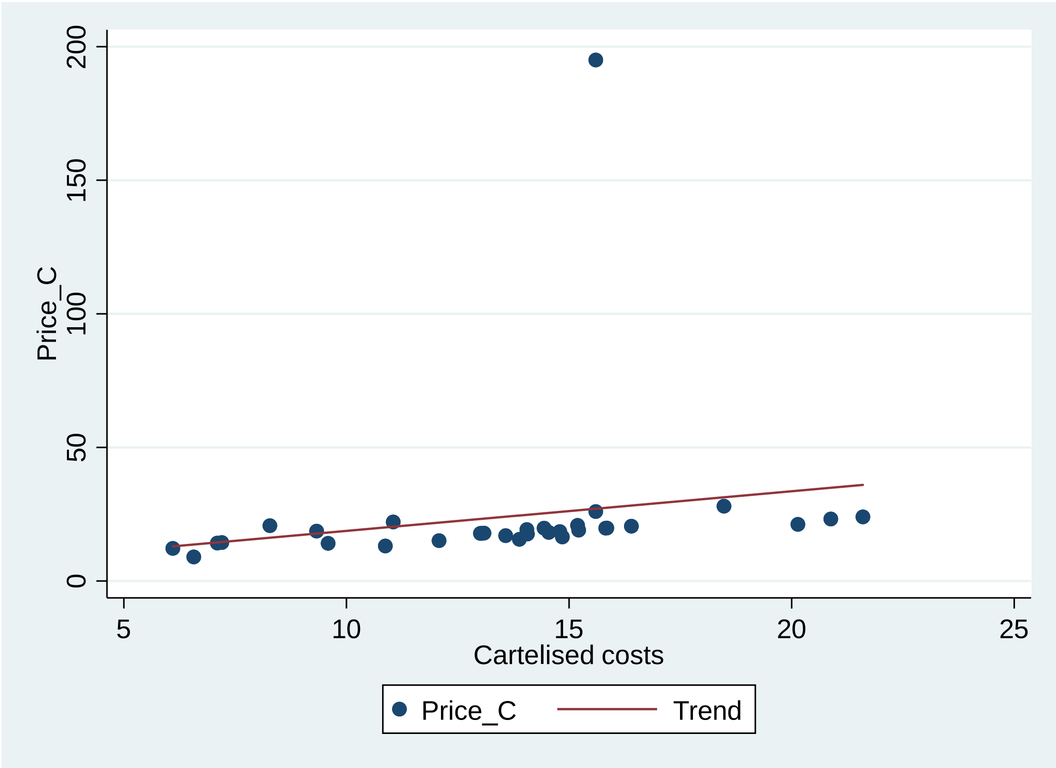
Statistic	Price_C	Price_C*
Mean	18.50 €	24.00 €
Median	18.60 €	18.60 €
Mode	19.80 €	19.80 €
Variance	15.6	989.6
Standard deviation	3.90 €	31.50 €
Coef. of variation	0.21	1.31
Q1	15.80 €	15.80 €
Q2	18.60 €	18.60 €
Q3	20.70 €	20.80 €
Maximun	28.00 €	195.00 €
Minimun	9.00 €	9.00 €
Range	19.00 €	186.00 €
Interquartile Range	4.80 €	5.00 €

Source: prepared in-house.

239. As illustrated in Table 3, statistics like the mean and variance are highly sensitive to outliers, indicating that the presence of outliers has a large impact on these statistics. The correlation coefficient between cartelised prices and costs is also strongly affected, from 0.77 to 0.19, simply because of the introduction of the outlier. The weakening of the previously robust and positive linear relationship between the two variables is evident in the following scatter diagram:

[125] This statistic measures the influence of each observation in an OLS regression, based on how the model output would change if that observation were omitted.

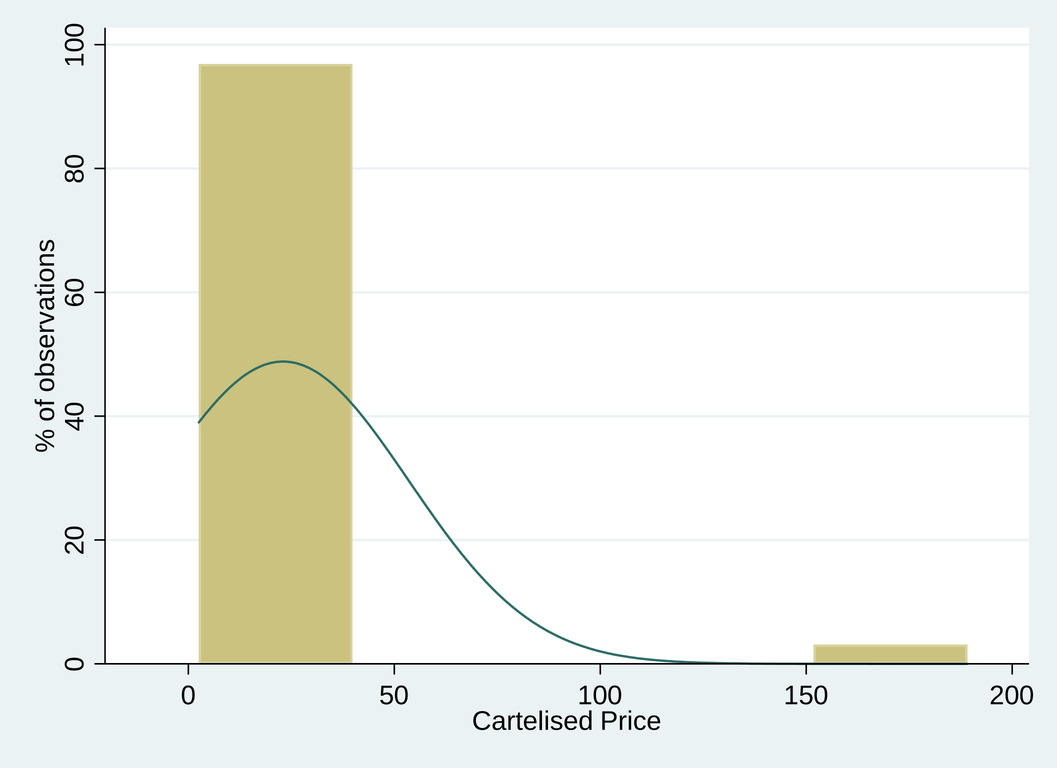
Figure 7. Scatterplot of price C and cost dispersion (including outlier).



Source: prepared in-house.

240. It is also possible to observe the sensitivity of the distribution to the presence of outliers through the cartelised price histogram, where the mere inclusion of an outlier turns a distribution with a high degree of symmetry into one with a strong rightward skew, as shown in the graph below.

Figure 8. Histogram of cartelised market prices



Source: prepared in-house.

241. In short, as illustrated through a simple example, it is essential to thoroughly analyse the data composition. This allows us to adequately describe the sample and to detect whether the analyses performed may be affected by the absence of certain observations or the presence of outliers[126]. This affects both the statistical and econometric analysis.

242. Finally, as far as the **treatment of outliers** is concerned, there are various options available (interpolation, deletion, correction, etc.) and the choice will depend mainly on the cause of the outlier (reporting error, missing values, belonging to different populations, etc.). In general, it is advisable to provide a a reasoned and transparent description of any data treatment, including a sensitivity analysis that shows estimates with and without treatment of outliers.

[126] After the introduction of the outlier in our example, it can be seen how the calculation of the overcharge through the comparison of average prices would vary significantly. Specifically, maintaining the atypical value of the sample, the overcharge would be 9.3 euros (24 - 14.7). However, if it is decided to eliminate this outlier and recalculate the mean or a more robust statistic for this phenomenon is used, such as the median, the overcharge would be significantly lower: 3.8 euros (18.5 - 14.7) or 5 euros (18.6 - 13.6), respectively.

A2.1.4 Statistical inference

243. **Statistical inference** is the set of techniques used to approximate the behaviour of a population based on information provided by a sample. The following two categories are usually distinguished.

244. On the one hand, **parameter estimation** involves obtaining information about the mean, variance and several other parameters from the sample. Since there is uncertainty in each estimate regarding the magnitude or sign of the parameter being estimated, one common practice for assessing the precision of an estimate is to construct a confidence interval[127]. A confidence interval is defined by two numbers within which the true value of a parameter is expected to lie with a certain probability. The higher the desired probability or degree of confidence, the wider the interval will be.

245. **Hypothesis testing**, on the other hand, aims to assess whether certain assumptions about the parameters of one or several populations are compatible with the information contained in the sample[128]. In any contrast, two hypotheses must be defined: a null hypothesis (H0), which in principle is assumed to be true and which includes the assumption that the parameter takes a given value, and an alternative hypothesis (H1), which is the opposite proposition[129].

246. In statistics, when testing a hypothesis, two types of errors can occur:

- **Type I error or false positive**: the null hypothesis, H0, is rejected when it is, in fact, true at the population level.

- **Type II or false negative error:** the H0 hypothesis is not rejected even though it is false.

247. Theoretically, the probability of making a type I error can be controlled by setting the desired **level of statistical significance**[130] (usually expressed

[127] For this, it is necessary to know the theoretical distribution of the parameter. The distribution is often assumed to be normal.

[128] For example, you may want to check if the average overcharge is equal to zero or if the average price in one market is higher than that in another.

[129] For example, if the null hypothesis (H0) states that the value of a parameter (e.g., overcharge) is equal to zero, its alternative hypothesis (H1) is that this value is not zero, which is known as a two-sided hypothesis test (“two-tailed”). However, there may also be "one-tailed" hypothesis tests, for example, if H0 assumes that the value of a parameter is greater than or less than a certain level (e.g., the overcharge is less than or equal to 10%), while H1 would indicate the opposite situation (continuing with the example, that the overcharge is greater than 10%).

[130] It must be taken into account that the smaller the type I error that one is willing to accept, the greater the probability of incurring a type II error and vice versa.

as a percentage; α%)[131]. When conducting a test, the **p-value** is usually calculated and is defined as the probability of obtaining a certain estimate assuming that the null hypothesis, H0, is true. If the p-value is lower than the chosen significance level (α%), the null hypothesis is rejected, and vice versa.

248. We can illustrate these concepts by using an example. Suppose one is analysing the relationship between the number of competitors (represented by the variable [$rival$]) and the price of a product (variable [P]) in a given market. In this case, it is assumed "by default" that no such relationship exists[132]. In other words, we adopt the null hypothesis, H0, that the coefficient of the explanatory variable to be tested (β, which indicates the effect of changes in the number of competitors on price) is equal to zero. Formally:

$$P = \alpha + \beta \cdot rival + \varepsilon \rightarrow H_0 : \beta = 0.$$

249. When testing H0 to decide whether to reject it or not, the concept of statistical significance level is used to reflect how stringent one is with the estimators obtained (estimated value of β using real data). The randomness of the observations leads to the estimators having a probability distribution around the true value of the coefficient. In principle, the more observations there are in the sample under study, the more centred the estimated coefficient will be on its true value[133].

250. Returning to the example of the number of competitors, let us assume that the p-value of our estimate is equal to 7.5%. We can interpret this p-value depending on whether we are more or less stringent in terms of the uncertainty of the result:

---

[131] As an example, a significance level of 5% implies that, if 100 different samples were randomly taken, on average, a type I error would be made 5 times.

[132] Generally, H0 is usually adopted in such a way that the value of the parameter to be tested is equal to zero, $H_0 : \beta = 0$.

[133] A small, unimportant effect may be statistically significant if there is a sample with enough observations to estimate, while a large effect may be insignificant if the sample size does not allow for an adequate estimate.

- Not rejecting $H_0: \beta = 0$ with a significance level of 5%, i.e., with a confidence level (probability) of 95%, we could not reject the hypothesis that the number of rivals has no relationship with the price of the product; or
- Rejecting $H_0: \beta = 0$ at a more permissive (less stringent) significance level, such as 10%, i.e., at a 90% confidence level we could reject the hypothesis that the number of rivals has a no relationship with the price.

251. Type II errors will be more unlikely the larger and more representative the sample analysed.

A2.1.5 Methods for comparing observations

252. The quantification of harm essentially consists of constructing a counterfactual and comparing it with the observed scenario. Several methods that may be useful for comparing various data sets are briefly presented below.

A2.1.5.1 Statistical tests

253. When comparing several samples, two types of statistical tests are usually distinguished:

- On the one hand, **parametric tests**[134], which assume knowledge of the distribution of the data and its main parameters: mean and variance.
- On the other hand, **non-parametric tests**[135], which do not CONSIDER assumptions about the population distribution.

254. Statistical tests can contribute, depending on the case, to an analysis and comparison of the factual and counterfactual scenarios. Parametric tests are the most commonly used, generally assuming that the variables are normally distributed. Their advantages include greater statistical power[136] and greater precision, provided that the underlying assumptions[137] are met. Non-parametric tests have the advantage of not requiring assumptions about distributions and being less sensitive to outliers than parametric tests.

[134] An example of a parametric test would be the Student's t-test (frequently used to compare means).

[135] Examples of non-parametric tests are the Mann-Whitney U test, Kruskall-Wallis H test, Wilcoxon test and Friedman test.

[136] With a parametric test, the probability of making a type II error is less than with an equivalent non-parametric test.

[137] For example, a parametric test could consist of comparing the means of two sets (such as the prices of companies that are members of a cartel and those of other non-cartel companies); this has the advantage of providing confidence intervals.

A2.1.5.2 The assumption of parallel trends

255. In Subsection 2.3.1.c of the Guide, the assumption of parallel trends (also called "parallel trends") was mentioned as a prerequisite for using the difference-in-differences method. Specifically, it is necessary to assume that the variable of interest through which the infringement is measured (e.g., price) would have evolved in the same way (in a "parallel" fashion) in the affected market and the reference market in the absence of the infringement.

256. The starting point for justifying that the parallel trend assumption is met is usually that the compared observations of the scenario affected by the infringement and the counterfactual were already evolving similarly before the infringement. This may require various types of analysis. On the one hand, a graphical analysis can be carried out to check whether the assumption is met. Caution is called for, as visual inspection may lead to very different conclusions depending on the length of the period considered or the scale used to construct the graphs. On the other hand, the parallel trend hypothesis can be tested using statistical and econometric techniques that examine whether there are significant differences in trends at points in time in the absence of infringements[138].

A2.2 ECONOMETRIC CONCEPTS

A2.2.1 General issues

257. In recent decades we have witnessed an increasingly frequent use of econometric techniques in various disciplines, including, among others, the quantification of harm for infringements of competition law[139].

258. The most frequently used techniques in the field at hand are regressions, which are used to try to understand and measure the relationships between two or more economic variables. In this case, the objective is to analyse the impact of anticompetitive conduct on the harm suffered by the plaintiffs. However, this task presents several difficulties:

[138] One possibility, when there are several periods before and after the treatment (the infringement), is to construct a binary variable for each period that interacts with the treated group (the one affected by the infringement). In order to consider that the supposition of parallel trends is fulfilled, the estimated coefficients of the previous periods should not be different from zero. See Subsection 4.3.1. from Annex 4 for a practical application of this technique. Another possibility is to use a placebo test performing the same analysis, but, for example, using a similar group to the treatment group, but which was not affected by the infringement, expecting that the results of these estimations are not significant.

[139] Its widespread use has been mainly the result of two circumstances. On the one hand, the technological development that allows the processing of large amounts of data in a very short time and, on the other hand, the theoretical development of industrial economics.

- Normally, the factor whose influence is to be quantified is not the only one affecting the variable of interest. As we have seen throughout this Guide, the key point is to **isolate the effect of an anticompetitive infringement from the rest of the variables** that simultaneously affect and determine the economic outcome.

- Moreover, even if the effects of other important systematic factors can be considered, there is a random error (or perturbation), since reality cannot be represented exactly in an equation (it will always remain a more or less close approximation).

259. Usually, regression models are used, which can be represented through the following generic equation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 \, x_2 + \cdots + \beta_k \, x_k + \varepsilon$$

260. Let us now analyse the different terms of the equation. On the left-hand side, we find the explained variable[140] (Y) which, in the field of harm quantification, is usually a variable used to measure the economic impact of the behaviour, such as price, profitability or benefits, among other possibilities. On the right side are:

- The explanatory variables[141] ($x_1, x_2 \ldots, x_k$), which are those factors[142] that influence the explained variable (e.g., supply and demand factors, regulations, the infringement);

- The parameters (β_1, ... , β_k), which measure the influence of each of the explanatory variables on the explained variable, keeping all other variables constant (*ceteris paribus*). The parameter β_0 (referred to as the intercept or constant) gives the predicted value of Y, should all other variables be equal to zero.

261. The error term (ε) (also called the disturbance term) captures unobserved factors that affect the dependent variable, and which are not directly explained by the independent variables[143]. The error term is unobservable and makes the relationship between the explained variable, Y, and the

140 It is sometimes called the dependent or endogenous variable.

141 They are also known as independent or exogenous variables.

142 In this general case, it is assumed that there are "k" explanatory variables. Typically, there is more than one variable, in which case it would be a multiple linear regression model (if there was only one, the regression model would be "simple").

143 The error term will depend mainly on the selection of variables, their treatment and potential measurement problems, amongst others.

explanatory variables, x_i, stochastic, i.e., subject to chance. As mentioned above, the usual objective of regression analysis is to detect the impact of changes in one or more explanatory variables on the explained variable. However, since the theoretical model that has just been presented is not observable, the aim is to obtain estimates of the parameters of interest that are as close as possible to their real value.

262. This requires: (i) a theory that indicates the variables to be used; (ii) the relevant data to be obtained; and (iii) the choice of an appropriate estimation technique and functional form[144]. In addition, to estimate the parameters of the model, several assumptions must be made, which are discussed in Section A2.2.3.

A2.2.2 How to choose the explanatory variables

263. Explanatory variables can be defined as those variables that, in addition to the anticompetitive infringement, may have influenced the dependent variable under analysis. These variables can be continuous (e.g., electricity consumption or raw material costs) or discrete (e.g., if the firm belongs to a certain cartelised region), and are modelled through dummy variables, which will be discussed in Subsection A2.2.2.2.

264. As an example, if raw material costs increased during the period of the infringement for reasons unrelated to the infringement (e.g., a bottleneck in the supply chain), the effect on the dependent variable could be wholly or partly the result of this increase. It is therefore important to separate, on the one hand, the exogenous (independent) effect of the raw material price increase and, on the other hand, the effect of the infringement. In such cases, the inclusion of an explanatory variable related to the cost of commodities in the model (e.g., CPI in the electricity subclass) would allow

[144] In this sense, there are several possibilities regarding the functional form of a regression model:

- Linear: $y = \beta_0 + \beta_1 x_1 + \varepsilon$, where β_1 indicates by how many units y changes if x_1 increases by one unit.
- Log-level: $\log(y) = \beta_0 + \beta_1 x_1 + \varepsilon$, where $(\beta_1 * 100)$ indicates the approximate percentage that y changes if x_1 increases by one unit.
- Level-Log: $y = \beta_0 + \beta_1 \log(x_1) + \varepsilon$, where $\frac{\beta_1}{100}$ indicates approximately how many units that y changes if x_1 increases by 1%.
- Log-Log: $\log(y) = \beta_0 + \beta_1 \log(x_1) + \varepsilon$, where β_1 indicates the approximate percentage that y changes if x_1 increases by 1%.
- Quadratic: $y = \beta_0 + \beta_1 x_1 + \beta_2 {x_1}^2 + \varepsilon$, where $\beta_1 + 2\beta_2 x_1$ indicates approximately how many units that y increases if x_1 increases by one unit.

its impact on the price to be considered and thus to capture more accurately the true impact of the infringement.

A2.2.2.1 Selection criteria

265. The inclusion of explanatory variables is intended to reflect factors not causally affected by the infringement and not controllable by the operators, but which may have been relevant for the dependent variable. In other words, explanatory variables must be exogenous, which also implies that these variables are uncorrelated with those variables not included in the model and thus absorbed by the error term. When selecting the explanatory variables, it is advisable to start from knowledge of the sector concerned, of the infringement and of economic theory[145].

266. The next step consists of a **joint relevance analysis** of the potential variables included according to the explanatory power they may contribute to the quantification, for which the following issues need to be considered:

    1. **Including too many explanatory variables** to reflect the same economic phenomenon might be an unwise practice, as multicollinearity problems may arise (see Section A2.2.5.3), which tends to lead to the overfitting problem of explanatory variables (e.g., R2).

       Suppose that the demand for the good or service upon which the infringement occurred is particularly sensitive to consumer income. In this case, one possibility would be to include the per capita income of consumers in the market. It might also be reasonable to include the unemployment rate if individuals tend to consume more when they are not unemployed. However, both variables, per capita income and unemployment rate, are strongly correlated, so that including both may be unnecessary and detract from the precision of the results. Therefore, the correlations between the potential explanatory variables should not be excessively high. In addition to analysing the correlation matrix, it is advisable to carry out other types of checks on the variables that generate the most doubts, such as sensitivity analyses based on various estimates using combinations of different variables.

    2. **Including variables that are not covered by the economic logic of the specific market**, based on a sufficiently high correlation with the dependent variable, is not a recommendable practice given that it may lead to the appearance of spurious relationships. In other words, it may be the case that a variable behaves very similarly to the dependent

[145] For a practical example of selecting explanatory variables, see Section A4.2.3..

variable without any causal relationship, simply because both are related to a third variable that is not considered. To avoid this problem arising, it is important to justify the inclusion of each explanatory variable and to include those that are relevant.

267. In conclusion, including explanatory variables is highly advisable when they are available, although it is important to select them **carefully based on adequate knowledge of the market** (with transparent and exhaustive arguments), which **could be complemented with quantitative analysis to assess their relevance in the model** (sensitivity analysis, calibration of the potential explanatory variables). In general, it is advisable to avoid the use of non-instrumented endogenous explanatory variables.

A2.2.2.2 Including dummy variables

268. A dummy variable is used to account for qualitative or discrete phenomena in a regression model (e.g., whether the firm belongs to a cartelised region or period or not) that normally takes values of zero (if it does not) or one (if it does). It could also be used to reflect the existence of several brands of the product being analysed (taking values equal to one for one brand and equal to zero for the rest), to capture the impact of a major regulation (by assigning zero to the periods before the regulation and one to the periods after), or to control for the seasonality of the data (by including a dummy variable for each period).

269. The coefficient of a dummy variable indicates the relative effect of belonging to a certain category (value one) with respect to belonging to the reference category (with only two categories, the reference category would be assigned a value of zero[146]). This is a fundamental difference with respect to the use of continuous explanatory variables, whose coefficients capture the impact of changes in the number of such variables on the dependent variable.

270. Two **types** of dummy variables can be distinguished:

1. **Additive:** additive dummy variables try to approximate the effect of a change in the group or category considered by the dummy variable, when this is assumed to be constant and independent of the value of the rest of the explanatory variables. For example, if it is considered that a company can manufacture a final product X using two alternative raw

[146] It is necessary to take into account that the number of dummy variables included must be equal to the number of existing categories minus one, in order to avoid perfect multicollinearity among the dummy variables included. Thus, if we wanted to capture the effect of four categories in a model, we would have to include three dummy variables and take one of the categories as a reference.

materials, A or B, one can try to isolate the relative effect on the price of X of using one raw material or the other, by including a dummy explanatory variable "Raw_mat_A" that takes a value equal to one in the case of using raw material "A" and equal to zero if raw material "B" is used. Thus, if the price is measured in euros and the coefficient of the dummy variable is estimated to be 0.05, it can be interpreted that, on average and keeping the value of the rest of the explanatory variables constant, using raw material A raises the price of the final product by 0.05 euros with respect to using raw material B.

2. **Multiplicative (interaction) variables**: this type of variable allows us to isolate the existence of simultaneous combined effects between two explanatory variables, when at least one of them is a dummy. Continuing with the example, let us assume that we have operator labour costs (a continuous variable, in euros) as an explanatory variable for price and that we suspect, from the economic logic of the case, that its impact depends on the raw material (the use of certain raw materials requires more labour). This relationship is modelled through the interaction (e.g., multiplication) of the two variables, creating a new variable. Thus, when using the model, estimated coefficients are obtained which provide the following information:

    i. Constant: the average price when the raw material is B (Raw_mat_A = 0) and the labour cost is zero.

    ii. Raw_mat_A: the variation in the price when the raw material is A (Raw_mat_A = 1) and the labour cost is zero.

    iii. Labour costs: change in price when the raw material is B (Raw_mat_A = 0) for each additional cent of labour cost.

    iv. Interaction variable: variation of the effect of labour costs on price when the raw material is changed from B to A. If it is positive (let us imagine 0.07), it means that the effect of labour costs is 0.07 points (in this case cents) higher for those operators using raw material A than for those using raw material B.

271. As discussed in the previous subsection, when using dummy variables, it is important both to **justify their use in the specific case** and **test their robustness** with the aforementioned analyses (e.g., showing models with and without interaction).

A2.2.3 The assumptions of the linear regression model

272. To estimate the parameters of the linear regression model (usually through the OLS method), it is necessary to make several assumptions, especially concerning the error term, since this term incorporates randomness into the model. The degree to which these are met will determine the properties of the estimators from the point of view of:

- centrality or *unbiasedness*[147]: the estimator, on average, is centred on the true value of the parameter,

- consistency: as the sample size increases, the estimates tend to get closer to their true value, and

- dispersion: it is desirable that its variability, depending on the sample chosen, be as small as possible (i.e., that it is efficient).

273. From a theoretical point of view, it is desirable that an estimator is efficient, consistent, and unbiased.

274. The assumptions of the linear regression model are as follows:

i. The model is linear in its parameters, i.e., the relationship between the variables in the model can be modelled as a straight line.

ii. The expected value of the error term is zero[148], so that no systematic errors are made when predicting Y.

iii. The error term does not correlate with the explanatory variables[149].

iv. Absence of perfect multicollinearity: none of the explanatory variables is constant or a linear combination of another explanatory variable[150].

v. The variance of the error term is independent of the value of the explanatory variables and is constant[151]. This is called homoscedasticity.

[147] An estimator is unbiased when its expectation is equal to the value of the parameter it intends to estimate.

[148] $E[\varepsilon] = 0.$

[149] $Cov(X_i, \varepsilon_i) = 0.$

[150] In this way, each explanatory variable contains additional information about the dependent variable that is not contained in the rest of the model.

[151] $E[\varepsilon_i - E(\varepsilon_i)]^2 = E[\varepsilon_i]^2 = \sigma_i^2.$

vi. The random terms are independent, there is no autocorrelation between the error terms of the different elements of the sample[152].

vii. The errors follow a normal distribution[153].

275. If the first four assumptions are met, the OLS estimators will be unbiased. Furthermore, if the conditions of homoscedasticity (v.) and absence of autocorrelation (vi.) are met, then the estimators will also be efficient.

A2.2.4 Analysing the regression results

A2.2.4.1 Statistical significance

276. When using econometric techniques for harm quantification, the debate often centres on the extent to which we can assume that the estimates of the regression parameters (β_0, ..., ...) inform us about its true value, i.e., about the true relationship between the explanatory variables and the explained variable. This is, therefore, a particular case of statistical inference in which the concepts introduced in Section 1.4. of this annex can be applied.

277. When presenting the results of the regressions, a common practice is to present the estimated value of each of the parameters, accompanied by its standard error, which measures how precise the estimation is[154]. When evaluating the estimated coefficients, it is necessary to consider three main issues:

- Sign: indicates whether the explanatory variable has a positive or negative influence on the explained variable.

- Magnitude: allows us to assess the significance of the effect.

- Significance: indicates the extent to which we can be confident that the effect is non-zero.

---

152 $Cov(\varepsilon_j, \varepsilon_i) = E[\varepsilon_j - E(\varepsilon_j)][\varepsilon_i - E(\varepsilon_i)] = E[\varepsilon_j \varepsilon_i] = 0.$

153 Studying the residuals of an estimate is important when validating the model. In theory, if the model adequately explains the relationships between the explained and explanatory variables, the residuals should be distributed approximately normally and independently with zero mean and a constant variance. The supposition of normality makes it possible to derive the probability distributions of the coefficients, β_i , and its variance.

154 In general, the higher the standard error, the lower the level of precision or reliability of the estimate.

278. Indeed, it is usual in econometric estimations to test the null hypothesis (H0)[155] that each of the parameters is equal to zero[156]. In other words, it is assumed that the different explanatory variables do not affect the explained variable and, if this is rejected, it is concluded that there is an effect. Depending on the case, some coefficients may be more important than others, especially when there is one that determines the value of the harm. Once again, statistical errors come into play, which in terms of quantifying harm means that it can be concluded that there has been harm when in reality there was none (false positive), or that there has not been harm when in fact there was (false negative)[157].

279. In this sense, the level of significance established is fundamental since it determines the degree of demand with which the results of the regressions are evaluated. In academic studies, the most widely used levels are 1%, 5% and 10%, implying that a probability greater than 99%, 95% or 90%, respectively, is required to consider that a parameter differs from zero. No threshold is preferable to another, it depends on the circumstances of the case (available evidence, presumptions, etc.), the quality and quantity of the data used, and so on; it is ultimately up to the judge to decide which threshold is acceptable.

280. The significance of the results is closely linked to the number of observations and the degree of collinearity (the higher the number of observations and the lower the multicollinearity, the more likely it is that statistically significant coefficients will be found). However, it should be borne in mind that some anticompetitive behaviour may have started a considerable time ago, making data collection difficult and leading to cases with a low number of observations. Thus, strict consideration of the significance level favours committing type II errors (false negatives).

A2.2.4.2 Goodness-of-fit

281. When assessing an econometric model, it is logical to ask to what extent the set of explanatory variables employed identifies changes in the explained variable. The most commonly used measure of the goodness-of-fit of a regression model is the **coefficient of determination**, **R2**. This indicator measures the proportion of the sample variation of the dependent

[155] For an illustration of a hypothesis test in the framework of a regression model, see Subsection 4.1.1 of Annex 4.

[156] To do this, the estimated coefficient is divided by its standard error and a ratio known as *t*-statistic is obtained. Another possibility is to test the null hypothesis that all the coefficients of a group of variables are equal to zero, in which case the *F*-statistic is used.

[157] This does not exclude the fact that there may be other types of errors such as under or overcompensation of real harm.

variable (Y) collectively explained by the independent variables of the model (X). Its value ranges from zero to one[158].

282. In general terms, it is preferable for R2 to be high. Otherwise, much of the variation of the dependent variable will depend on factors not included in the model (omitted variables), so that its explanatory power will be reduced. Otherwise, much of the variation of the dependent variable will depend on factors not included in the model (omitted variables), so that its explanatory power will be reduced. However, caution is required when interpreting the results of the coefficient of determination for several reasons:

- A high R2 reflects the fact that there is a high degree of correlation between the explanatory variables and the explained variable, but this does not imply that there is a causal relationship. Therefore, a causal relationship may be adequately estimated with a low R2.
- The value of the coefficient may depend on the characteristics of the analysed data, including the sample size, the time dimension[159], the level of aggregation of the variables, or the functional form of the dependent variable.
- If the number of variables in a model is increased, it is likely that R2 will increase[160]. This may encourage the inclusion of a very large number of variables to achieve a high R2, even when the marginal contribution of each of the new variables is not statistically significant. To alleviate this problem, the adjusted **R2** is created, which will only increase with the inclusion of an additional independent variable if it adds some explanatory power to the model, otherwise its value will decrease[161]. However, with large samples, the difference between the two coefficients tends to be diluted.

283. In general terms, it can be stated that there is no value at which the R2 of a model is considered sufficiently high (low) to be able to validate (discard) it. The main objective of a model used for harm quantification should not be to maximise the adjusted R2, but rather for the model to have been constructed on the basis of reasonable assumptions from the point of view

---

[158] An R2 equal to zero would imply that none of the independent variables help to explain the variation of the dependent variable, while if it is equal to 1, it means that the explanatory variables perfectly capture such variation.

[159] It often happens that, in time series models, the R2 is higher than with cross-sectional data, simply because the variables present common trends.

[160] The inclusion of an additional variable may not change the value of the coefficient if its explanatory power is zero, but it will never decrease.

[161] Specifically, it will only increase if its *t*-statistic is greater than one (in absolute value). In extreme cases, its value could become negative.

of economic theory and to be able to estimate the causal effect of the explanatory variable of interest on the dependent variable as well as possible.

284. Note that it is not possible to compare the R2 of models with different specifications or estimated using different methods. However, if we consider the same model, it may be interesting to study the joint significance of all or a group of explanatory variables, for which the F-statistic, which is closely related to the coefficient of determination[162], is often used.

A2.2.4.3 The sensitivity of the results

285. All econometric estimation has a degree of uncertainty associated with it in terms of the validity of the chosen functional form, the estimation method, the variables selected, the data used, and so on. A sensitivity analysis allows us to see how changes in the assumptions of an economic model affect its results and, in this way, can help to validate its results and provide a range of possible estimates. In principle, we would expect the main conclusions of the model to remain unchanged despite changes in certain assumptions. However, it is important to design this analysis properly, otherwise it could be used to reinforce models and conclusions that are originally flawed.

286. The first step in a sensitivity analysis is to decide which assumptions you want to test. There is a wide range of issues that can be tested, and this depends on each particular case. The idea is not to modify every assumption, but only those that may be the most controversial. For example, one can **reasonably exclude various control variables** and test how the coefficients of the explanatory variables of interest are affected. One can try to exclude variables individually or jointly, notably if they are considered interrelated. It may also be possible to include **several periods throughout the infringement**, if there are doubts, to modify the **functional form** of a variable, or to reasonably exclude certain **outliers** from the sample and check for changes in the predictions. For reasons of transparency, it is always advisable for expert reports to reflect those points in the analysis to which the model presented is most sensitive.

[162] In general, a high value of R2 will correspond to a high value of F, which implies that the set of independent variables explains the variations of the dependent variable. However, it may happen that the hypothesis contrast using the F-statistic suggests that there is joint significance of the variables, while R2 presents a low value. In the latter case, we have a statistically significant model, but the explanatory power of the independent variables is low. For a practical application of both concepts, see Subsection A4.4.1.2.

287. A sensitivity analysis can also suggest changes to the econometric model. For example, if there are uncertainties between various specifications of a model (different variables, functional forms, etc.), it is possible to compare these using statistical tools such as adjusted R2 or F-statistics and decide accordingly[163].

A2.2.5 Frequent problems

288. Sometimes the regressions proposed do not meet one or more of the assumptions of the "classical" linear regression model, explained in Section A2.3. The most common problems of econometric estimations and their possible solutions are described below, highlighting the implications for harm quantification. Although these problems are present in the estimates, it is necessary to assess, among other issues, the relevance of the problems detected, their magnitude, the solutions adopted and the alternatives available, before ruling out these models.

A2.2.5.1 Functional specification error

289. One of the assumptions of the linear regression model is that there is linearity in the parameters ($\beta_0, \beta_1, \beta_2,..., \beta_k$). However, it should be borne in mind that this is a relatively flexible assumption since it does not require the relationship between the explained variable and the explanatory variables to be linear[164]. In fact, it is quite common for certain variables in the models to be expressed in logarithmic, quadratic or exponential forms to try to capture non-linear relationships, without this invalidating the estimation of a linear regression model; however, it does change the interpretation of the coefficients, so it is necessary to take this into account.

290. It is therefore important to construct econometric models properly, bearing in mind that economic theory does not usually determine the functional form of the relationships between economic variables. Failure to adequately represent the relationships between variables[165] leads to an error in the functional specification, which results in biased and inconsistent estimators.

[163] The main difference between the two is that, while with the adjusted R2 models it is possible to compare models with different specifications (in principle, those with a higher adjusted coefficient would be chosen), to compare two models using the F-statistic, it is necessary for one of the models to be a particular case of the other with fewer explanatory variables, in order to be able to contrast their joint significance.

[164] An example of a linear relationship would be if the increase in one cost always gave rise to the same increase in price, regardless of the levels of both variables.

[165] For example, assuming that the relationship between price and energy costs is linear, when in fact it is logarithmic.

291. To detect functional specification problems, one possibility is to visually inspect the relationships between the variables in advance using scatter plots to show which type of trend line (e.g., linear or quadratic) best summarises the relationship between the observations of the two variables. One can also use the Ramsey RESET test[166] or add transformations of variables that may be misspecified and assess their joint significance with respect to the baseline model using the F-test.

A2.2.5.2 Endogeneity

292. In econometric terms, endogeneity can be defined as the existence of a correlation between an explanatory variable and the error term. This violates one of the assumptions of the linear regression model and prevents an approximation of the individualised effect of the explanatory variables on the explained variable, giving rise to biased OLS estimators (differing on average from their true value) and inconsistent estimators (no matter how much the sample is increased, they will not approach their true value).

293. This phenomenon arises from having elements included in the error term that are related to explanatory variables in the model. As these elements are not included in the model, it can be complex to detect the endogeneity problem, so it is always advisable to start from economic reasoning and knowledge of the market analysed to assess the possible existence of endogeneity.

294. This problem can be the result of several circumstances including measurement errors[167], autoregression[168] with autocorrelation of errors, simultaneity[169] or omitted variables[170]. The solutions will depend on the origin of the problem (the deduction of which is mainly based on economic theory or knowledge of the sector). If it arises from the omission of relevant variables, one solution would be to include these variables directly in the

[166] The test includes non-linear combinations of the explanatory variables and verifies whether they contribute to explaining the dependent variable (if so, the model would be poorly specified).

[167] A measurement error may appear, for example, as a consequence of deficiencies in data collection or due to erroneous aggregations of certain variables.

[168] In autoregression models, the variable explained at the current time (t) is influenced by its past (at time t-1, or even by earlier times). For this reason, it is necessary to include different lags of the explained variable as an explanatory variable. These models in econometrics are called AR(p), where p indicates the number of periods that go back in time (lags) to identify the effects of the past on the present.

[169] Simultaneity occurs when the independent variables of one model appear as dependent variables in other equations and vice versa. An example would be if one tries to explain the price of a product with its quantity demanded by means of a single equation, since, although an increase in demand can affect the price, changes in price will also have an effect on demand.

[170] An omitted variable is one that is not included in the model as an explanatory variable but, nevertheless, influences the dependent variable or other explanatory variables (that is why it is in the error).

model or, if this is not possible[171], to use proxy variables, which correlate with them[172]. If this is not possible, one could also attempt to justify the direction of the bias in the estimator of interest and indicate whether it can be expected to be upward or downward (i.e., overestimated or underestimated, respectively)[173].

295. Another possible solution to the endogeneity problem is the approximation and substitution of the endogenous variable (the one that causes the problems) with instrumental variables. An instrumental variable is a variable that does not belong to the model (it is independent of the error term) and correlates with the endogenous explanatory variable[174]. While the first condition (exogeneity) is not observable[175], the second (relevance of the instrument) is, and it is desirable that the correlation between the instrumental and endogenous variable be as high as possible.

296. When one or more instruments are available, the model is usually estimated using the two-stage least squares method (2SLS). In the first stage, the endogenous explanatory variable is regressed on the instruments ("auxiliary regression") and the rest of the exogenous explanatory variables, checking the relevance of the instruments by analysing the significance of their coefficients. In the second stage, the original model is estimated using OLS, the only difference being that the actual values of the endogenous variables are replaced by the values predicted in the first stage.

297. If both above-mentioned conditions are met, the 2SLS estimators will be consistent. Nevertheless, if there is no actual endogeneity problem, OLS estimators are preferable as they exhibit greater efficiency[176]. For this purpose, one can apply, among other things, the Durbin-Wu-Hausman test, which compares the OLS and 2SLS estimators and assesses whether their

[171] For example, due to the absence of data or the impossibility of obtaining information.

[172] For example, a possible approximation to the evolution of the labour costs of a company would be to observe the variation of the unit labour costs of the sector to which it belongs.

[173] However, when there are several explanatory variables that act in the opposite direction, predicting the direction of the bias is more complex.

[174] For example, if we want to estimate the demand for a product and we include price as an explanatory variable, an endogeneity problem can be expected to arise (for example, unobserved factors that affect the demand for the product such as perception of quality; it can also affect the price consumers are willing to pay). A possible solution would be to apply the evolution of the cost of an input used in the manufacturing of the product, since it can be expected to positively affect the price (relevant) but it is not likely to affect the final demand for the product (exogenous).

[175] It must be based on economic theory or some other supposition, so one has to be very careful when choosing an instrumental variable.

[176] The standard errors will always be higher in the case of an estimation using 2SLS rather than OLS, which implies greater efficiency in the second case. However, in case of endogeneity, the OLS estimation will not be consistent.

differences are statistically significant[177]. For an example of how to deal with the endogeneity problem, see Subsection A4.4.1.3.

A2.2.5.3 Multicollinearity

298. When it is indicated that a model presents problems of multicollinearity, it means that the correlation between some explanatory variables is high[178]. This problem can occur with some frequency in harm quantification, when strongly related supply or demand control variables are included. For example, if different cost variables are included in the model, it is possible that they are positively correlated and that, although together they have a positive influence on price, it is difficult to discern the individual effect due to multicollinearity problems.

299. In the presence of imperfect multicollinearity, OLS estimators will still be unbiased and consistent, but not efficient. This implies that the standard errors of the estimators will be larger, so that the estimation loses precision and certain variables may be considered non-significant, when in fact they are.

300. However, this is not usually one of the most serious problems, insofar as it only affects the precision of certain estimators, while unreliability and consistency will not be affected. Thus, if in the harm quantification we are only interested in the coefficient of a certain explanatory variable to detect the effect of an infringement, it will not be a problem if other control variables present multicollinearity (at the cost of losing some efficiency). However, if the variable of interest is affected by this problem, it may be difficult to identify the effect we are interested in capturing.

301. To detect the existence of imperfect multicollinearity, it is useful to calculate the correlation matrix between the explanatory variables[179]. Another

---

[177] The underlying logic is that, in the absence of endogeneity, both estimators are consistent, so they should give similar results. Therefore, if there is a significant difference, it may indicate that there are endogenous variables.

[178] This phenomenon is called "imperfect" multicollinearity. In practice, it is not possible for perfect multicollinearity to occur, since it would be impossible to obtain estimates of the parameters. Therefore, if an explanatory variable turns out to be a linear combination of others, econometric packages automatically detect its presence and suppress the problematic variable.

[179] The higher the values (closer to 1 or -1, depending on whether the correlation is positive or negative, respectively), the more likely it is that there is multicollinearity.

frequently used statistic to analyse whether multicollinearity affects a particular variable is the variance inflation factor (VIF)[180].

302. The best solution to the problem of inefficiency (larger standard errors) caused by multicollinearity is to try to increase the sample size to reduce the standard errors of the affected parameters. Another option is to transform or eliminate the variables that are the most problematic, provided that excluding them from the model makes economic sense and does not foreseeable introduce a new endogeneity problem. If the multicollinearity is not very clear and does not affect the variables of interest, it may be preferable not to adjust the model at all.

A2.2.5.4 Heteroscedasticity

303. Earlier we assumed that the model was homoscedastic, i.e., the error term had constant variance across the observations and over time; otherwise, the model presents problems of heteroscedasticity.

304. Heteroscedasticity is more frequent with cross-sectional data, especially when the units analysed (individuals, companies) do not behave homogeneously. It can occur for various reasons, including samples constructed from aggregating individual data, outliers (especially in small samples), poor model specification, or the structure of the data itself.

305. As with multicollinearity, heteroscedasticity means that OLS estimators are no longer efficient, although they are still unbiased and consistent.

306. To analyse whether the model exhibits heteroscedasticity, it is usual to start with a graphic analysis of the residuals, comparing them with the predicted dependent variable and the independent variables[181], it being desirable to obtain a random structure, free of trends. Another tool consists of plotting the observed values against the predicted values and comparing them with the unit slope line, i.e., 45° (they should be close to this slope). After a graphic exploration, the analysis can be reinforced by statistical tests[182].

307. The problem of heteroscedasticity can be addressed in several ways:

[180] The higher the value, the greater the indication that there is multicollinearity. Sometimes the limit is set at 10, but this is still an arbitrary value, so a decision on a model should not be made just because the VIF is high. For more information, see Subsection A4.4.1.2.

[181] This can be used to identify the variable that is furthest from randomness as the cause of the problem.

[182] Among others, there would be the contrasts of White, Goldfeld-Quandt and Breusch-Pagan.

- Solving model specification problems: one can resort to changing the functional form[183], dealing with outliers, excluding the exogenous variables that are causing the problem[184], and so on.

- The most common solution, when heteroscedasticity is suspected and its form is not known, is to use standard errors that are robust in terms of heteroscedasticity[185], although a large sample size is required.

- Another possibility (less used in practice) when the form of the heteroscedasticity is known is to use GLS[186] estimation, rather than OLS.

A2.2.5.5 Autocorrelation

308. When the assumption that errors are independent is violated, there is a problem of autocorrelation. Autocorrelation generally occurs in time series, so that errors in one period influence errors in subsequent periods. This is particularly relevant for comparative methods using multi-period data. For example, if a shock not included in the model increases prices in one period by more than predicted, it is possible that the error will remain positive in neighbouring periods. On the other hand, it is also possible for this problem to appear in a cross-sectional sample, and it is common if regional economic data is available, since the economic situation of several regions may be affected by the same shocks.

309. The most frequent causes of autocorrelation are:

- The existence of cycles or trends in the dependent variable that are not reflected in the model.

- A misspecification of the model, by choosing the wrong functional form or omitting variables that are correlated over time.

310. As with heteroscedasticity, autocorrelation means that OLS estimators are no longer efficient and statistical inference is also affected.

---

[183] A common transformation to reduce the heteroscedasticity problem or to facilitate the interpretation of the results in percentage terms is to express some variables in logarithms.

[184] However, this can in turn generate an omitted variable bias, so it is necessary to be guided by economic theory and the characteristics of each case.

[185] Known as Eicker-Huber-White standard errors.

[186] Specifically, the method of Weighted Least Squares is usually used, which gives less weight to the observations with greater variance in the error term.

311. Graphic methods[187] and hypothesis testing[188] can be used to detect autocorrelation.

312. There are several ways to solve this problem. On the one hand, if the autocorrelation stems from specification problems, one can carry out variable transformations or try to include the explanatory variables into the model that were omitted, always assuming that these are justified from an economic perspective. Alternatively, the model could be estimated using GLS instead of OLS. Standard errors that are robust in terms of autocorrelation can also be used.

A2.2.6 Particularities of data with a time dimension

313. Frequently, the data used in harm quantification has a temporal dimension, in different formats[189]. This gives rise to a series of peculiarities, some of which have already been mentioned throughout this annex. Other issues are highlighted below, such as adjustments that may need to be made to the data prior to processing or estimation methods specific to the data panels.

A2.2.6.1 Possible data adjustments

314. It is relatively common for certain economic variables to rise over time, displaying a more or less common **trend**. Therefore, when quantifying harm, one can attempt to take into account the effect caused by a trend, to avoid attributing the effect to another explanatory variable. This can be particularly useful in cases where the evidence suggests that there is a trend in the variable of interest that holds throughout the data series and cannot be explained by the rest of the explanatory variables[190].

315. The first step is to analyse whether any of the variables included in the econometric model have a trend and then try to capture this as best as

[187] Usually, autocorrelation functions (simple and partial) are used, which relate a variable with the same variable in previous periods, to find the autocorrelation level of the data.

[188] Among others would be the contrasts of Durbin-Watson (the most common), Wallis, Breusch-Godfrey and Box-Pierce.

[189] Time series, pool, or panel data.

[190] Instead, if there appear to be one-off shocks, it might make more sense to add time dummy variables rather than tendencies.

possible[191]. Once the trend is recognised, it should be included as an explanatory variable in the model to avoid omitted variable[192] bias.

316. However, it should be borne in mind that including a trend variable in a model can significantly impact the outcome of the harm quantification. Therefore, good practice involves adequately justifying the inclusion of the trend, its functional form and carrying out a sensitivity analysis to show that it is not the key variable that generates or masks the harm.

317. Another problem that may arise with time series when they are presented at a frequency greater than annual (quarterly, monthly, weekly, etc.) is **seasonality**. In the event that any of the variables used present seasonal behaviour, it may be necessary to make certain adjustments (known as "deseasonalising")[193]. The logic is the same as with the trend: to avoid attributing effects deriving from the time of year considered to certain variables[194]. Although there are various methods for factoring in the seasonality of the data, some of which are highly complex, a relatively simple option is to include dummy variables in the econometric regression according to the period[195] to which each observation corresponds and to analyse whether these are significant.

318. Finally, it should be borne in mind that the variables to be included in the analysis may have **different periodicities**, which makes it necessary to carry out transformations so that all the data has the same periodicity. For example, if some variables have a monthly periodicity and others are quarterly, there are several viable options: (i) omit the variables with the lowest frequency (quarterly) from the analysis; (ii) aggregate the variables with the highest frequency (i.e., convert monthly variables into quarterly ones); (iii) perform the analysis at the highest frequency level (monthly),

[191] To do this, it will be necessary to take into account which function best approximates its evolution over time: linear, quadratic, exponential, etc.

[192] For example, $Y = \beta_0 + \beta_1 X_1 + \beta_2 t + \varepsilon$, where we would expect β_2 to be positive (negative) if Y increases (decreases) over time (t) for reasons unrelated to X_1.

[193] On occasions, the data series have already been previously seasonally adjusted.

[194] For example, certain agricultural or construction activities are influenced by the weather, which will vary depending on the time of year.

[195] For example, certain agricultural or construction activities are influenced by the weather, which will vary depending on the time of year. It will be necessary to include one variable less than the periods to avoid multicollinearity problems.

using an imputation technique to substitute the values of the variables with the lowest frequency (quarterly)[196].

A2.2.6.2 Lagged variables

319. In data with a time dimension, the explanatory variables of a model sometimes include the dependent variable, lagged by one period. An example would be to try to explain the price in one period using the price in the previous period, as a way of reflecting the existence of rigidities in the price adjustment or simply to try to include factors that influence the price which are not reflected in the rest of the explanatory variables.

320. The problem with including a lagged dependent variable is that, in the case of an autocorrelation issue, several problems may arise:

- The estimators of the coefficients of the explanatory variables become biased and inconsistent.

- Both the significance of the coefficient of the lagged variable and the R2 of the model are likely to become artificially high, while the rest of the variables lose significance.

321. To prevent the lagged variable from having excessive weight in the model, various measures can be taken, such as extending the frequency of observations (e.g., using quarterly instead of monthly data) or taking the first differences of all the variables[197].

A2.2.6.3 Panel data estimation methods

322. By combining cross-sectional and time-series information, panel data allows us to control for the unobservable heterogeneity of the agents studied, i.e., intrinsic characteristics that do not vary over time and are relevant for explaining the dependent variable[198]. In the absence of such data, this heterogeneity would be reflected in the error term, giving rise to potential endogeneity problems. Depending on the assumptions made about the nature of these unobservable effects, different estimation methods can be applied.

---

196 All options will have their advantages and disadvantages. Notably, options (i) and (ii) involve giving up some of the available information, while option (iii) involves making assumptions about the behaviour of the missing information that may be debatable and affect the quantification result.

197 In the models expressed in first differences, all the variables are transformed by subtracting the same variable from the immediately preceding period.

198 If the agents are companies, these unobservable characteristics that can influence the price (or another dependent variable) would be intangible, such as the quality of the products, the brand image, etc.

323. If we assume that the unobservable effect correlates with an explanatory variable in the model[199], two methods are usually applied to perform the estimation: (i) the **first differences** method or (ii) the **fixed effects**[200] method. On the other hand, if we assume that the unobservable heterogeneity does not correlate with the rest of the explanatory variables, a **random effects** model will have to be applied.

324. Depending on the circumstances of each case, it may be preferable to use fixed or random effects estimators. To decide which of the two to use, the Hausman test[201] is usually applied.

[199] For example, if R&D spending is included as an explanatory variable and the productivity of each company (assuming that it cannot be measured) is positively correlated with it.

[200] To estimate a model using fixed effects, you can include a different dummy variable for each unit analysed or transform each (dependent and explanatory) variable, by subtracting its time average.

[201] This test is based on the hypothesis that the fundamental assumption of the random effects method (null correlation between unobservable heterogeneity and explanatory variables) is fulfilled. If it is rejected, it means that the estimates of fixed and random effects are significantly different, so it is preferable to use the fixed effects method.

ANNEX 3: REVIEW OF THE ECONOMIC LITERATURE

325. A review of the relevant economic literature is always important to contextualise the state of the issue, as well as to obtain examples of the main practices used. As mentioned above, the estimation of harm is a unique exercise that should avoid the mechanical application of estimation percentage ranges applied in other cases. Therefore, this annex, far from attempting to be exhaustive, focuses on citing examples and relevant **methodological considerations** that **support the other messages contained in this Guide** and offer the reader the opportunity to **expand their knowledge**, especially regarding the practical application of the various quantification methods.

326. Under these premises, there is a multitude of publications related to the quantification of harm in the context of competition law infringement, with the majority of them focused on cartel cases, notably in the literature with origin in the United States. This annex analyses certain examples because of their particularly informative or explanatory nature, without prejudice to the existence of many other publications of the same kind.

A3.1 THEORETICAL AND METHODOLOGICAL STUDIES

327. Numerous theoretical studies deal with the quantification of harm from the perspective of economic theory. They sometimes include empirical sections, but these are hypothetical and not based on real cases. They generally deal with topics such as:

1. The economic theory of harm (**Baker and Rubinfeld**, 1999; **Motta**, 2004; **Rubinfeld**, 2008; **Davis and Garcés**, 2009; **Lyons**, 2009; **Maier-Rigaud and Schwalbe**, 2013; and **Niels**, 2016).

2. The theoretical and practical framework for analysing cost pass-on defence (**Hellwig**, 2006; **Kosicki and Cahill**, 2006; **Davis and Garcés,** 2009 o **Verboven and Van Dijk**, 2009).

3. Competitor foreclosure is the focus of the analysis of **Fumagalli, Padilla and Polo** (2010), who highlight additional difficulties with respect to cartel infringements due to dynamic effects on markets. The theoretical framework is precisely illustrated through different phases (attrition, recovery, and reactivation) that require individualised study. Along the same lines, but in greater detail, are the guidelines by **Fumagalli, Motta and Calcagno** (2018).

4. The increased use of econometric models (and their usefulness) in follow-on cases in Europe (**Droukopoulos, Veronese and Witte**, 2020). The authors argue in favour of using regression analysis which,

although it may seem unintelligible to non-specialists, can increase the accuracy of a harm estimate, thus helping to achieve a higher standard of evidence if the applicable regulations so require. Although regression analysis adds complexity to the study, its advantages are highlighted, such as the simultaneous treatment of several factors (demand, prices, product characteristics, costs, macroeconomic and other exogenous variables), the limitation of uncertainty, and the treatment of how entry and exit of competitors affect prices.

5. The need for several assumptions and caveats in the use of econometric models when calculating the cost pass-on rate in harm estimation cases (**Harris and O'Sullivan**, 1979). In the same vein, it is emphasised that the key point of econometric analyses is to isolate the effect of an anticompetitive infringement from the other conjunctural variables and to demonstrate causality between the infringement and the economic outcome (**McFadden et al.**, 2003).

6. The need to maintain a balance between pragmatism and precision in the development and presentation of harm quantification methodologies, emphasising values such as transparency and clarity, and seeking a meeting point between legal and economic professionals (**Friederiszick and Roller**, 2010).

7. The practical application of the main methods for quantifying harm using simulated data (**Heller and Maier-Rigaud**, 2021).

8. The main methodological considerations of difference-in-differences analyses (**Maier-Rigaud and Sudaric, 2019**)

9. The relevance of presenting different types of specifications according to different levels of statistical significance for the sake of greater estimation transparency, without the need to use the levels typical in other types of work (**Johnson et al.** 2017). In this sense, the work of **Bönisch and Inderst** (2019, 2021) proposes the concept of "severity" to support decision-making in legal proceedings when the parties present contradictory statistical evidence, expanding the range of options beyond accepting or rejecting estimates based on either their significance or averaging across results.

A3.2 EMPIRICAL STUDIES

328. Below are some of the publications that address the analysis of specific cases, highlighting the most important messages related to the methodologies used.

A3.2.1 Publications with methodological comparisons applied to specific benchmark cases

329. **Finkelstein and Levenbach** (1983), **Rubinfeld and Steiner** (1983), and **Fisher** (1980, 1986) analyse how to apply econometric techniques in procedures for claiming damages for infringements of competition law, through various real cases in the United States.

330. **Daggett and Freedman** (1984) critically analyse the evidence presented in a cartel formed by the US canned tomato industry over the period 1951-1975. They outline the step-by-step construction of an econometric model, including easy-to-understand explanations of the level of significance and error, and make several specific recommendations, already mentioned throughout this Guide:

    1. They detail the baseline market situation and consistently describe the infringement, which may also include arrangements for purchase price reductions from suppliers. This background information is crucial for the proposed estimate.

    2. They evaluate the adaptation of the model to reality, showing the range of options available or ruled out.

    3. They adjust the cost variables to reflect the effect of inflation over time.

331. In addition, **Harrington** (2004), uses an analysis of harm quantification in the US graphite electrodes cartel (1992-1997) to emphasise how important it is to consider whether there was a time lag before market conditions returned to the pre-infringement situation following the termination of the anticompetitive conduct. Moreover, when the impact of the infringement cannot be clearly separated in time from other circumstances, because, for example, the beginning or end point of the infringement is not known with certainty, it is appropriate to omit periods that give rise to doubt.

332. Also noteworthy is the work of **Friederiszick and Roller** (2010) on the lessons learned from the critique of the expert reports submitted in Germany for the cement cartel and the paper wholesale cartel at the end of the 20th century. It is stressed that the approach adopted by the courts consists of three phases: design, implementation and robustness checks. In the design phase, it is reasoned that approaches based on regional comparisons (the cartel was too widespread in the rest of the German regions and probably in neighbouring countries), and market comparisons (no similarities were found) should be excluded. As a result, a comparison of different time periods or diachronic comparison approach was chosen, limited to the period during and after the cartel, with considerations on the relevance of

data aggregation and price wars in the design of the method. However, although the analysis seems to comply with the recommended caveats, the courts lowered the amount of the quantification proposed by the experts. Given this situation, the authors provide a detailed and technical account of the difficult balance between pragmatism and technical rigour in the valuation of expert opinions, highlighting the need for a common framework of understanding and knowledge between jurists and economists.

333. **Notaro** (2013) applied a variety of harm quantification methods to the 2007 pasta cartel in Italy. The author highlights that, in general, econometric methods (such as a binary variable approach with dynamic treatment effects) perform better than simpler methods, whose results tend to be particularly biased when there have been significant changes in demand or costs over the course of the infringement. Finally, the paper reiterates the need to correctly determine the level of penalties for anticompetitive practices as a deterrent factor and the enormous economic impact of the competition authorities' interventions.

334. **Connor** (2014b) analyses the estimates presented in the framework of the lysine amino acid cartel, which ran from 1992 to 1995 in the United States. In particular, he critically analyses the five most commonly used methods (market comparison, diachronic, difference-in-differences, cost-based and structural methods) and underlines the possible heterogeneity of the results depending on the approach chosen and the assumptions made, as well as the need to take into account the global dimension of the cartel when setting the amount of the damages, to prevent jurisdictional fragmentation from undermining the deterrence factor of the compensation.

335. In the same vein, **Seixas and Lucinda** (2019) analyse the Brazilian hydrogen peroxide cartel (1995-2004) to show the broad spread of estimates for harm that can result depending on the model applied. By means of examples they emphasise the need for proper justification for the use of the models. and offer several alternatives that enhance the credibility of the estimates. It is also highlighted that, in order to select the correct time period, other factors should be taken into account such as the cyclical fluctuations of the economy, any significant changes (shocks) in the markets, as well as the temporal delimitation for the beginning and end of the anticompetitive conduct.

A3.2.2 Publications based on comparative methods

336. Siotis and Martinez-Granado (2010) quantify the harm caused by the incumbent operator in the Spanish telephone information services market for hindering the entry of new operators by increasing costs after the liberalisation of the market in 2003. Based on what happened at the same time in a similar situation in the British market and using econometric tools (geographical comparison)[202], they approximate the market share that the new entrant would have had in the absence of t**he infringement.**

337. **Vanssay and Erutku** (2011), use the petrol station cartel that existed in Sherbrooke (Canada; 2000-2006) to compare the evolution of petrol prices in Sherbrooke and Montreal (geographical comparison).

338. **Boswijk, Bun and Schinkel** (2019) demonstrate, at both a theoretical and empirical level, the importance of clearly delimiting the temporal duration of an infringement. Based on the example of the sodium chlorate cartel in Europe (1994-2000), they estimate that using the legal duration of the cartel instead of the effective duration results in a 25% lower quantification of harm.

339. Turning to the combination of the above comparative approaches, the **difference-in-differences method** has particularly attracted the interest of researchers, since the number of publications has been quite high in recent years.

340. **Hüschelrath et al.** (2013) use the cement cartel in Germany (1991-2002) to illustrate the fundamental nature of temporal delimitation for the diachronic models and difference-in-differences, particularly in relation to the possibility of transition periods that have a crucial impact on the resulting estimates.

341. **McCluer and Starr** (2013) use a real case of harm quantification in the health sector in the United States to illustrate the advantages and potential disadvantages of using this methodology.

342. Furthermore, **Laitenberger and Smuda** (2015) focus on the harm suffered by German consumers caused by the washing powder cartel in Europe 2002-2005, and offer an estimate that combines the diachronic model, to assess the existence and magnitude of umbrella effects in other products, together with the difference-in-differences method to calculate the overcharge. Throughout this publication, there is emphasis on the fact that once the database has been constructed, and especially if it contains sources of diverse origin, homogenisation and processing is required paying special attention to the handling of data with the same level of aggregation both from the point of view of the timeframe (annual, monthly, weekly, daily

[202] Another analysis of this case can be found in Hitchings (2010).

or hourly data) and the differentiation of the product based on its characteristics.

A3.2.3 Publications on cost-based and financial methods

343. Even though courts frequently resort to costs as a basis for calculating harm when they are not convinced by the other methods presented or in the absence of quality data, **there are not many specific publications in the literature**. In addition to the works already mentioned that deal with this methodology along with others, the work of **Veljanovski** (2019) on the cartel in the submarine electrical wiring tender that operated between 1999 and 2009 (BritNed case) stands out. The author questions the decision of the magistrates regarding the interpretation of the information on direct costs, the calculation of gross margins, and the compensation factors related to cost savings generated by the cartel.

A3.2.4 Publications based on structural models

344. Structural models are often used as a framework for obtaining estimates of passing-on costs. The study by **Cotterill and Dhar** (2003) analyses the passing-on at the different stages of transformation of the liquid milk market in Boston (United States) over the period 1996-2000 (Nash and Stackelberg vertical models). In addition, **Kim and Cotterill** (2008) propose different estimates of demand and market structure (Nash-Bertrand equilibrium, collusion, etc.) to estimate the impact of costs (especially variations in milk costs) in the US processed cheese industry.

A3.2.5 Publications on the application of interest

345. **Gotanda and Sénéchal** (2009), focusing on the case of arbitration proceedings, argue that the compensation granted by the courts is usually insufficient as it does not take into account the time value of money and instead references risk-free investment interests, which business agents rarely undertake. The authors collect various possible interest rates and argue in favour of those based on the opportunity cost of capital and calculated in a compound manner. **Dow** (2022) works on the framework of international arbitration and presents different rates and forms of capitalisation, together with the advantages and disadvantages of each one.

346. **Bueren et al.** (2016) compare how different jurisdictions (United States, England and Wales, France and Germany) take into account interest and inflation in damages claims resulting from f competition law infringements. Likewise, the authors use a real example (the lysine cartel that occurred in the United States over the period 1992-1995) to simulate the economic impact of the different approaches presented, giving rise to hypothetical quantifications that can be almost three times higher in some jurisdictions

than others. The article highlights the relevance of three factors: (i) the point in time at which the interest begins to be applied, (ii) the magnitude at the national level of the interest rate applied before and after the judgement, and (iii) whether the interest is applied as compound interest.

A3.3 SYSTEMATIC REVIEW AND META-ANALYSIS

347. Finally, there are publications that review, for different practices in different industries, other studies (systematic review and meta-analysis[203]) and offer a series of recommendations, highlighting the following:

348. **Connor and Bolotova** (2006) review more than 800 estimates of price overcharge caused by cartels that have occurred between the 18th century and the beginning of the 21st century in the United States, Canada, Europe, Australia and Asia. They conclude that the longest-lasting cartels, with international dimension and characterised by their high concentration, tend to cause greater harm, while a downward trend in the amounts of harm is observed when the competition authorities increase their control over the cartels. Similar findings have been found in successive reviews by **Connor** (in 2008, 2010 and 2014), in which, for the same geographic area as in the first study, a considerable increase in harm estimates can be seen (with more than 1,200 new estimates since 2004) as a result of the increase in remedies imposed by the competition authorities. Furthermore, these publications emphasize, from a general perspective, the importance of selecting a suitable methodology tailored to the specific characteristics of the case and data availability. **Oxera** (2009), after making a series of adjustments to the data provided by Connor and Lande (2008), analysed the overpricing of 114 cartels, highlighting the importance of paying attention to the distribution of the data and not only to the mean or median, but also the need to delve into the specificities of each case. Finally, the analysis by **Bolotova** (2009), complements the previous conclusions by pointing out that cartels with many participants and those with unequal market shares among them tend to cause lower overcharges.

349. In line with the estimates in the reviews by Connor and Bolotova, **Smuda** (2012) analyses the level of overcharges on a sample of 191 cartels in the European market to detect the factors that can explain regional differences in the magnitude of the overcharge. The conclusion is that the overcharge is higher in cartels involving international firms than in those involving domestic ones, that participation in public procurement has a positive effect

[203] A meta-analysis is a systematic review of the studies carried out and the results obtained using a statistical tool that allows the results of these studies to be added together and analyses the existence, or not, of a relationship between them (Castellanos and Solano, 2017).

on the overcharge, indicating potential signs of collusion, while the effect of duration may be ambiguous.

350. Based on the database compiled by Connor (2010), **Boyer and Kotchoni** (2015) critically review the cartels included therein, concluding that estimates of overcharges above 50% are more likely to be biased. Furthermore, diachronic and synchronous comparative methodologies, as well as cases where there was a price war, tend to obtain higher estimates than those based on costs, econometrics, legal decisions or merely theoretical methods.

ANNEX 4: A PRACTICAL EXAMPLE

INTRODUCTION

351. The purpose of this annex is to use a **practical example** to illustrate several of the methods presented in this Guide (focusing on the comparative techniques, as these are the most common), showing some of the statistical and econometric techniques for the preparation of expert reports for the quantification of harm due to anticompetitive conducts and, in this way, facilitate its subsequent evaluation. It is important to emphasise that the inclusion of certain methods and techniques in the practical example **does not imply that these are considered preferable to other options not covered**. Additionally, there is no intention to rank the methods and techniques in the practical example since their selection depends on the availability of data and the specificities of each case.

352. The examples presented have been constructed using a **simulated database** and are intended to highlight the careful treatment of methodologies that, while not exhaustive or mandatory, is desirable when quantifying harm. This practical and schematic example is intended to introduce, in a simple way, econometric concepts that are particularly relevant when analysing expert reports and thus promoting good practices in those reports.

353. The structure of the example is as follows. First, the infringement (in this case is a cartel) is described in terms of the actors involved, the time frame of damages and the selection of variables. Second, the descriptive statistics of the relevant variables are presented together with figures that facilitate a clear understanding of the distribution of the observations that are under analysis. Thirdly, the methods used to quantify the overcharge are presented:

   i. A **synchronous method**, which compares the prices of cartel and non-cartel companies during the infringement period.

   ii. Two **diachronic methods**, which use data from the cartel companies from different time periods:

      - The first, a **dummy variable approach**, compares prices in the period affected by the infringement with prices in the periods before and after the infringement.

      - The second, a **predictive approach**, is based on the pre-cartel and post-cartel periods.

   iii. A **difference-in-differences** method.

354. Finally, the harm quantified with the different methods is capitalised and the example is concluded highlighting that the estimates obtained are complementary and fall within a range depending on the methodology and assumptions adopted.

A4.2 DESCRIPTION OF THE CASE

355. This example focuses on the analysis of an **intermediate product** (direct purchasers use it to produce a final consumer product), which is homogeneous and is produced in two regions of the same country, A and B, in each of which there are five factories. For the sake of simplicity, it is assumed that **the factories produce only this product**, so that it is not necessary to analyse the allocation of costs between the different business branches (as would be necessary in the case of a "multi-product" company).

356. In addition, it is assumed that the competition authority of the country in question has sanctioned the infringement of competition rules for the price-fixing of the intermediate product between the five plants in region A. The period is limited to January 2012-December 2013 and this sanction is firm, so it would imply a follow-on claim. Specifically, it was found that the five factories had agreed on the product prices to be charged to their customers (direct purchasers). Therefore, in this case, the anticompetitive conduct mainly affects these direct buyers of the intermediate product, who claim compensation for the potential harm suffered, which will focus solely on the calculation of the overcharge and interest[204].

A4.2.1 Timeframe of the infringement

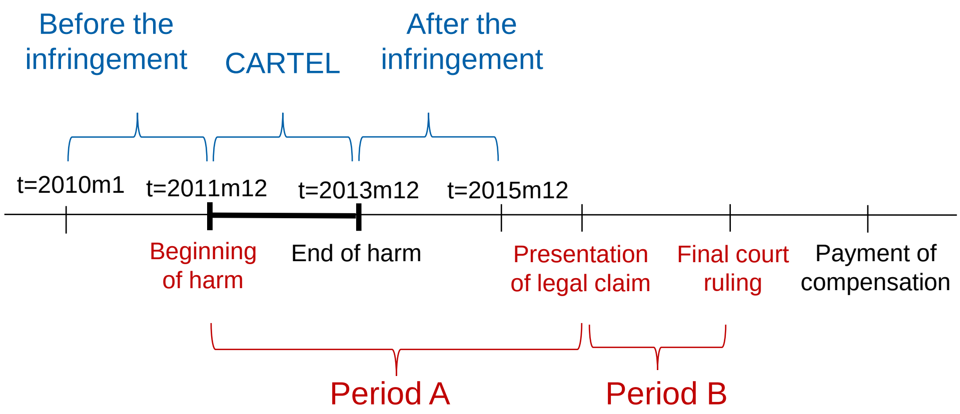
357. To quantify the damage, **monthly data are available for six full years (2010-2015)**, that is, a total of 72 months. To simplify the analysis, it is assumed that the duration of the infringement fully coincides exactly with the duration of the potential damage to the direct buyers. In this way, there are no lagged effects and **there was only overcharging during the existence of the cartel, which took place during the third and fourth years under consideration**, in accordance with the decision of the competition authority sanctioning the infringement.

358. Thus, from the point of view of calculating the price premium, the timeframe of the practical example can be divided into three periods: before, during and after the infringement, which is shown in blue in Figure 9.

[204] For simplicity, it is assumed that all the companies in the market in which the infringement occurs participate in it, so it is not necessary to assess possible umbrella effects.

359. It is assumed that **a direct purchaser of the cartelised product claims compensation** for the harm incurred during the two years of the cartel. The lower part of the figure shows the legal milestones in red (the beginning of the damage, the filing of the lawsuit and the date of the ruling of the first instance) that will ultimately affect the capitalisation of the damage, as will be shown later.

Figure 9. Timeframe of events related to the infringement



Source: prepared in-house.

A4.2.2 Description of the counterfactual

360. As has been pointed out in Subsection 2.2.2 of the Guide, knowledge of the harm (in this case, the overcharge derived from price rigging between manufacturers) is the basis for building the counterfactual (the situation that would have existed if the infringement had not taken place).

361. Throughout the example, **different counterfactuals** are presented depending on the quantification methods used. Although this may not be feasible in a real case due to limited economic resources, time, or data availability, it has the advantage of allowing us to present the analysis of the same hypothetical case from different angles and to explore how the results of the harm quantification might vary.

362. On the one hand, there is a **synchronous model** ("market comparison") whose counterfactual is made up of factories with similar characteristics to the factories in Region A, but these have not been affected by the infringement. The cornerstone of this model is to be able to justify the similarity between factories with and without the infringement, possibly through the use of relevant explanatory variables. In this way, five factories in Region B are included and the aim is to capture the differences with the factories in region A, considering a series of **control variables** (cost of

materials, price of electricity, labour costs and regional gross domestic product), which will be detailed in the following section.

363. On the other hand, a **diachronic model** (“comparison of different time periods”) is constructed using data from Region A itself in periods not affected by the infringement as a counterfactual. Two models are presented: one based on dummy variables (to capture the overcharge by means of a **dummy variable** that takes the value of "1" in the infringement period and "0" in the non-infringement period), and another based on a **predictive approach** (predicting price developments during the infringement period based on data from the non-infringement period).

364. Finally, the **difference-in-differences method combines the synchronous and diachronic methods** and compares Regions A and B (first difference), at times affected and not affected by the infringement (second difference).

A4.2.3 Selection of the relevant variables and the data used.

365. The first step is to determine **variables** to identify and quantify the damage. In this specific case, as it is a price-fixing cartel, it is considered that the most direct variable for quantification is **the price paid by direct buyers for the cartelised product**. In particular, the main harm comes from overcharging, that is, the difference between the prices actually paid by direct buyers during the infringement period in Region A and the counterfactual prices, which approximate the price they would have paid in the absence of the cartel. For the sake of simplicity, any reference to the effects of the pass- and volume effect is omitted.

366. The next step is to select the **relevant variables** that have an effect via supply and demand on the evolution of the prices during the months observed. In this way, the explanatory variables are selected according to the underlying economic theory and knowledge of the sector concerned. They are then presented, with a description of their role in determining the price, and they are grouped according to whether they affect the supply or the demand side[205].

367. It should be noted that variables have a maximum of three dimensions (they are shown with a subscript): (i) *f*, which indicates a factory, (ii) *r*, referring to

[205] It should be noted that the names of the variables are irrelevant and that they are only of interest for the purpose of illustrating the quantification methods and techniques used.

regions A or B, and (iii) *t*, the period under observation. A balanced data panel is used, containing observations of each factory in all months.

368. On the supply side, several **cost variables** are considered:

a. The **cost of materials** [$C_{mat_{ft}}$] reflects the cost of supplying all types of materials (expressed in euros per unit of product) necessary for the production process. It is assumed that individualised data is available at the factory level. The cost of materials has a strong (positive) direct relationship with the price of the product (see the coefficients in Table 5), which is also **quadratic**, implying that their relationship is not constant but varies depending on the level of costs considered:

b. It is assumed that the production of the product is intensive in the use of electric energy. In the absence of individualised data on the consumption of electrical energy in the factories, **the wholesale price of electricity** [P_{enel_t}] is used to approximate the monthly evolution of the cost referring to this source of energy. This variable is common to all factories in both Regions A and B and has a time dimension with a monthly frequency. It should be noted that these monthly data are the result of an aggregation of (average) daily data to adjust their frequency to that of the other variables used in the model.

c. **Labour costs** are included in the analysis because they are considered a relevant component of the variable costs of the factories. It is assumed that individualised labour cost data by factory are not available, so a *proxy* variable is used: an **index of monthly labour cost** in each Region A and B [$Indlab_{rt}$] published by the official statistical office of the country.

369. On the **demand** side, **regional gross domestic product** (GDP) [PIB_{rt}] is used, which captures, at an annual frequency[206], the general evolution of economic activity in each of the Regions A and B. It is assumed that there is a positive relationship between the price of the product and GDP, so that when the economy expands, demand grows, and prices increase. Prices tend to rise, and the opposite happens when GDP falls.

[206] Since the frequency of this variable is lower than that of the rest, the database uses the same value of GDP for each region in the months of the same year. This does not preclude the use of other techniques to deal with variables with different periodicity. (see Subsection A2.2.6.1).

A4.3 STATISTICAL ANALYSIS

370. In any quantitative analysis, including for the quantification of harm in competition law infringements, it is desirable to **present descriptive statistics** of the variables used throughout the analysis, to **analyse the structure of the data and to demonstrate the transparency of any processing carried out**. The following is a non-exhaustive presentation of some techniques that could be used[207].

371. Table 4 presents various descriptive statistics for the variables described previously, distinguishing between Region A (with five cartelised factories) and Region B (with five non-cartelised factories). The statistics show that the biggest difference between these two regions is found in the price variable, while the rest of the variables show similar values in all the statistics analysed[208]. The labour cost index is slightly higher in Region A than in B, which could help explain the apparent higher price.

Table 4. Descriptive statistics of the variables used

Variables	Unit	region A				region B			
		Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
price	€	27.66	14.27	1.79	60.25	20.59	8.64	1.67	44.91
cost of materials	€	16.03	0.45	0.47	2.47	16.06	0.43	0.48	2.54
wholesale price of electricity	€/MWh	47.98	0.96	2.38	6.29	47.98	0.96	2.38	6.29
labour cost index	index	104.76	2.41	96.22	107.84	100.82	2.21	92.68	102.98
regional GDP	mil €	0.49	0.02	0.46	0.51	0.46	0.03	0.42	0.50

Source: prepared in-house.

372. The analysis then focuses on the variable of interest (the price) that will be used to quantify the damage, the explanatory variables and the correlation between the variables.

Variable of interest

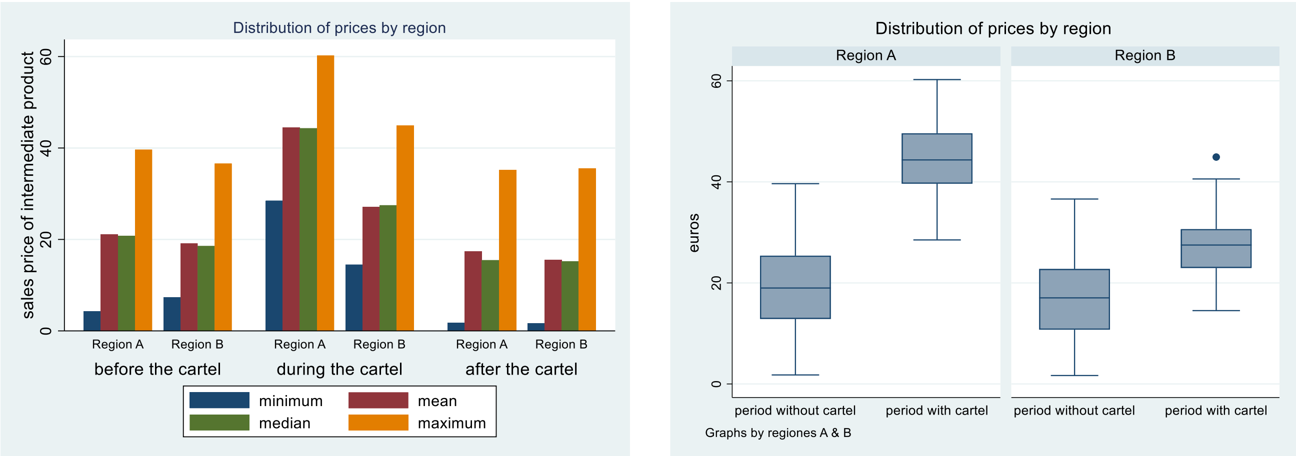
373. Figure 10 shows that price differentials between the regions occur mainly during the existence of the cartel, with similar behaviour in the preceding and subsequent periods. On the left, each set of columns represents descriptive statistics (mean, median, minimum and maximum) referring to the cartelised or non-cartelised market in each of the three periods

207 The descriptive analysis of a real case may not be as evident as in this annex, due to the simplifying assumptions used and the simulated nature of the data. This would be the case, for example, in the presence of non-homogeneous products or different supply or demand conditions.

208 As a complement, it is pointed out that the variable with the greatest difference in the dispersion of the observations between regions is the price, which in Region A presents a standard deviation of 14.3 euros, while in Region B it is 8.6 euros.

considered[209]. The four statistics used to illustrate the dispersion and centrality of the data, show that prices in both regions were slightly higher in the earlier period than in the later period. On the right is the distribution of prices, using a box-and-whisker plot, such that the non-infringement period includes both the pre-infringement and post-infringement periods. Although a higher price is observed in both regions during the infringement period, the difference is greater in the cartelised region.

Figure 10. The price of the product by region and period



Source: prepared in-house.

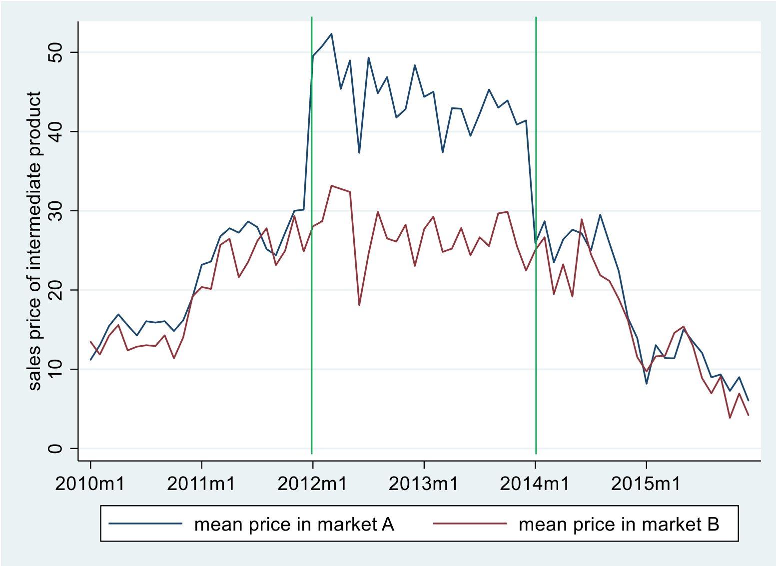
374. Once the main magnitudes of the variables under analysis have been described, a preliminary assessment of the behaviour of the variable of interest (price) over time is made. This is done by plotting the evolution over time of the average monthly prices of the product in the different markets. Figure 11 shows a similar evolution of prices in the periods before and after the cartel in the factories[210]. On the other hand, the divergence during the infringement period is particularly important, denoting a possible upward influence on prices by the cartel[211].

[209] For more information on these metrics, see Subsection 1.2.1 of Annex 2.

[210] It should be noted that this parallel evolution, also referred to in the relevant literature as "parallel trends", is a necessary condition to make differences-in-differences type estimates, discussed in more detail in Section A4.4.3, where a quantitative verification of said condition is also presented.

[211] The difference is probably excessively clear compared to what would be observed in a real case. Again, this can be attributed to the use of simulated data to facilitate illustration of the techniques and methods.

Figure 11. Evolution of average monthly prices in markets A (cartelised) and B (non-cartelised)



Source: prepared in-house.

375. In addition to the graphical analysis, the average prices over the three periods are compared by employing a t-test. In particular, the following pairwise comparisons are carried out:

1. For region A: the prices of the product are compared (i) before and during the cartel, and (ii) during and after the cartel.

2. Between regions A and B: the prices of the product are compared (i) before, (ii) during, and (iii) after the cartel.

376. The null hypothesis (H_0) of these contrasts is that there is no difference between the means of the pairs of mean prices. We adopt a significance level of 95%, which means that we will reject H_0 (i.e., we would find that there are differences in means) if the contrast is associated with a p-value of less than 5% (or 0.05). We get two results:

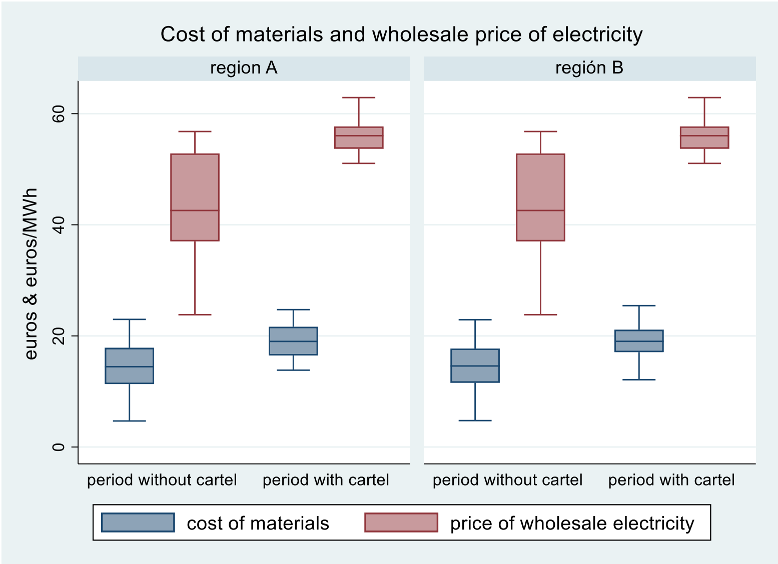
- On the one hand, the result of these contrasts shows a p-value of zero for the comparisons of region A, in other words, **the data suggest that the prices during the cartel were higher on average than the prices in the periods not affected by the infringement**.

- On the other hand, the average prices in markets A and B are only significantly different during the cartel (p-value of zero), while in the periods before and after the cartel it cannot be excluded that they are the same (p-value greater than 0.05)[212].

Explanatory variables

377. Additionally, to identify the effect of explanatory variables on price, it is important to understand the structure and distribution of the data. For this, graphical visualisation is a useful tool, for example, using box-and-whisker plots (Figure 12)[213].

Figure 12. Distribution of certain explanatory variables



Source: prepared in-house.

378. On the supply side, it is assumed that the costs of materials and electricity are the main determinants in product prices, due to the characteristics of the production process. While the costs of materials are closely linked to the technology implemented and the use of inputs (e.g., raw material) in each factory, the price of electrical energy is an external and aggregate variable, representing wholesale prices of electricity in the country. Figure 12 shows that the levels of the two variables relating to costs of materials and electricity prices were lower in the two regions in periods outside of the infringement compared to the periods during the infringement.

[212] For more details from a theoretical perspective, see Section A2.1.4 of Annex 2.

[213] For other illustrations of box-and-whisker plots, see Subsection A2.1.2.1 ("Measures of dispersion") in Annex 2.

Correlation analysis

379. Apart from describing the different variables separately, as a preliminary analysis it can be useful to examine the correlations that exist between them and with the variable to be explained (price). Table 5 shows the average correlation coefficients between the different variables, where a positive or direct correlation is observed between price and cost variables, especially in the case of material costs and the price of electricity. Similarly, there seems to be a slightly positive relationship between price and regional GDP[214].

Table 5. Matrix of correlation coefficients between the variables

Variables	price	cost of materials	wholesale price of electricity	labour cost index	regional GDP
price	1				
cost of materials	0.81	1.00			
wholesale price of electricity	0.78	0.75	1.00		
labour cost index	0.21	0.03	0.06	1.00	
regional GDP	0.23	0.10	0.19	0.34	1.00

Source: prepared in-house.

A.4.4 ECONOMETRIC ANALYSIS

380. The econometric analyses presented below are framed within the typology of reduced-form comparative methods. These methods condense the relationship between the variable of interest (in this case, the price of the intermediate product) and a set of explanatory variables into a single one, which tries to capture demand and supply factors that are believed to influence the former[215].

381. Concerning the variables included in an econometric estimation, it should be noted that it is useful to have a perception *a priori*, based on knowledge of the sector analysed, economic theory and previous analysis of available data, of at least two aspects:

[214] However, as indicated in Annex 2 (Subsection A2.1.2.2), it is not a good idea to draw definitive conclusions from simple correlations.

[215] For more information on these methods, see Subsection 2.3.1 of the Guide.

1. **the importance of each explanatory variable for the evolution of the price**, which is reflected in the magnitude of each coefficient estimated in the regression models,

2. **the direction of the effect of the change in the explanatory variable on the dependent variable** or, in other words, the sign of the estimated coefficient. The relationship is direct or positive (positive coefficient) when the variation in the explanatory variable is associated with a variation in the same direction in the dependent variable, and the relationship is inverse or negative (negative coefficient) when a variation in the explanatory variable is associated with a change in the opposite direction in the dependent variable.

382. In this annex, as mentioned in the introduction, three quantification methods are presented (1. Synchronous, 2. Diachronic, with dummy variable and predictive approaches, and 3. Difference-in-differences) to see through an example their main advantages and disadvantages, as well as possible difficulties that may arise in practice. In addition, it is expected that the results obtained with the different methods will not be same, but rather oscillate within a range. The following table provides a schematic representation of the three methods applied and their distinctive characteristics, such as the periods and markets considered, as well as the variable that captures the overcharge.

Table 6. Diagram of the quantification methods applied

Variable of interest: PRICE	1. Synchronous method	2. Diachronic method: a. Approach with dummy variable	2. Diachronic method: b. Predictive approach	3. DID
Periods considered	during	before/after vs during		
Markets (regions) considered	A & B	A	A	A & B
Variable that identifies the overprice	$cartel_f$ *	$period_t$ *	observed price - prediction	DID_{ft} ** = $cartel_f \cdot period_t$

Source: prepared in-house.
Note: (*) dummy variable, (**) interaction of dummy variables.

A4.4.1 Analysis comparing different markets during the infringement period (Synchronous comparison)

383. The method is based on the comparison of the prices of the product in Region A with those of Region B during the infringement period, since it is assumed that the development of Region B was competitive and can be

used as a reference to approximate the non-cartel situation in market A[216]. Thus, there are 240 observations: the 24 months of the infringement multiplied by 10 companies (5 in market A and 5 in market B).

384. First, a simple regression model (with a single explanatory variable) is presented to introduce some basic concepts. Then, multiple variables are introduced to solve a potential endogeneity problem by using instrumental variables.

A4.4.1.1 Simple regression model

385. We start by presenting the regression model [S_simple] with a single dummy explanatory variable, $\boldsymbol{cartel_f}$, which equals one for the observations corresponding to cartelised factories (region A) and zero for the non-cartelised factories in region B. The following simple regression with the dummy variable $\boldsymbol{cartel_f}$ is estimated using OLS and the results are presented, below, with the regression equation.

[S_simple] $$P_{ft} = \beta_0 + \beta_1 cartel_f + \varepsilon_{ft}$$

$$\widehat{P_{ft}} = 27.1048 + 17.3626 \cdot cartel_f$$

386. The **constant term** [β_0] indicates the price level if the explanatory variable ("cartel") were equal to zero[217]. In this case, in the absence of the cartel, the intermediate product would cost 27.10 euros. The estimate of the parameter β_1 indicates the **average overcharge** associated with the cartel. In other words, according to the model data, during the infringement period, cartelised factories (those for which the variable $\boldsymbol{cartel_f}$ takes the value of 1) they sell the intermediate product at an average price that is 17.36 euros higher than the price applied by a non-cartelised factory. Both results are obtained with a confidence level of 99%[218].

---

216 Another option for a synchronous analysis would be to compare similar products rather than similar geographical markets.

217 It should be noted that the interpretation of the constant term in multiple regressions is more complex from an economic point of view, so it is not usually a relevant part of the methodological discussion.

218 The second column of Table 7 of results indicates with (***) that the p-values of the estimated parameters are less than 0.01 and, therefore, with a confidence level of 99%, the null hypothesis can be rejected, that their population values are equal to zero.

Table 7. Results of the estimation of the simple regression model

Dependent variable: price of intermediate product		S_simple
cartel	beta	17.3626 ***
	standard error	0.7974
	t stat	21.77
	p-value	0.0000
constant	beta	27.1048 ***
	standard error	0.5638
	t stat	48.07
	p-value	0.0000
Number of observations		240
F test		474.14
Prob>F		0
R2		0.6658
R2_adj		0.6644

Legend: * level of confidence interval 90%; ** level of confidence interval 95%; *** level of confidence interval 99%

Source: prepared in-house.

387. It is important to note that the estimated parameter value may differ from its true value. The measure of this deviation is the standard error, with high values indicating a wide range of possible outcomes and low values indicating a more precise estimate (with less uncertainty).[219,,220].

388. Another measure of the significance level for the estimate of a parameter is its **t-statistic** (third row of each variable in Table 7), which is the quotient of the estimated parameter and its standard error (being $H_0: \beta_1 = 0$). The higher the value of this statistic (in absolute terms), the stronger the evidence against the null hypothesis, which in this case points to the existence of overcharging. Another alternative way of analysing the significance of an estimate is to use the ***p-value*** (fourth row of each variable in Table 7), which provides information about the probability of obtaining a particular estimate assuming that the null hypothesis, H_0, is true. In other words, the smaller the *p-value* of a coefficient, the more certain we can be that its true value is not zero.

389. In this example, the estimated standard error is very small relative to the estimated parameter value. Therefore, the probability that the null

[219] In this case we have standard errors of the estimators β_0 and β_1, respectively, 0.5638 and 0.7974.

[220] Generally, a 95% confidence interval is established, which is calculated as the estimated value of the parameter plus/minus two times the standard error (in our case, it would be $17.36 \mp 2 * 0.79 = [15.765; 18.955]$.

hypothesis β_1 is equal to zero ($H_0: \beta_1 = 0$) can be rejected is greater than 99%, corresponding to the observed p-value of 0. In addition, the coefficient of determination indicates a goodness-of-fit level equal to 0.59. Thus, belonging to the cartel explains 59% of the variations of the price over the periods analysed.

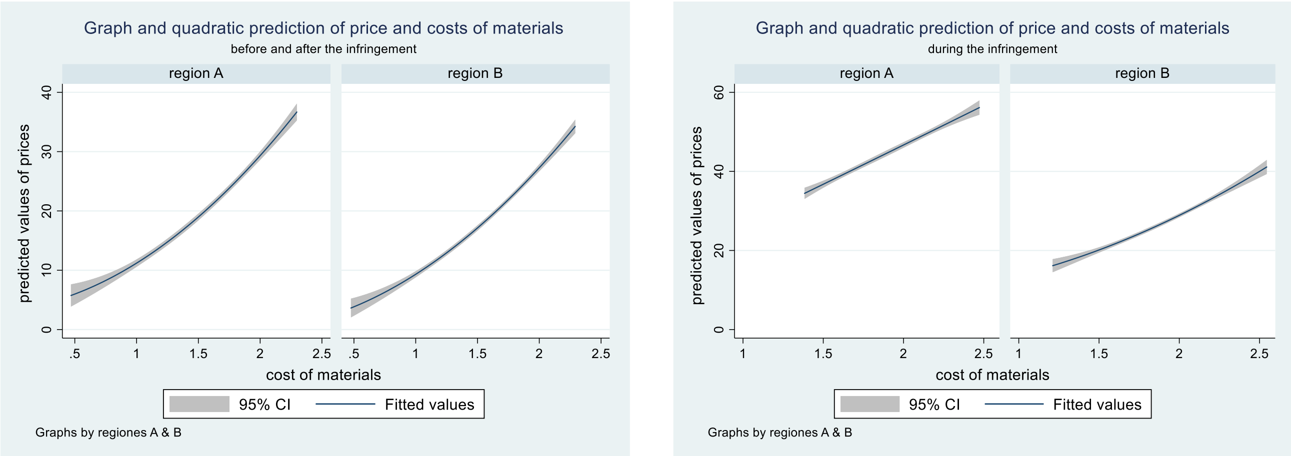
A4.4.1.2 Multiple regression model

390. In the following multiple regression model, [S_multiple], in addition to the dummy variable "cartel", we include the available variables considered relevant (explanatory variables) that may affect the price in the market affected by the infringement, since, otherwise, these variables would be included in the error term, and could affect the quantification by introducing a bias in the coefficient of the infringement variable.

391. As we have already indicated, for the **supply side**, we include the variables cost of materials, wholesale price of electricity and labour cost index. Meanwhile, we approximate the **demand side** by using the regional GDP variable.

392. A fundamental question is to **determine the functional form of the regression** (e.g., linear, quadratic, logarithmic, exponential, etc.), that is, the representation of the relationship between the dependent variable and each of the explanatory variables. Based on the knowledge of the sector, the existence of a **quadratic relationship** between the material cost variable and price was detected, and **to illustrate this,** a figure is presented showing this quadratic adjustment in different periods. In this case, a quadratic function reflects that price growth is greater as the value of the explanatory variable (cost of materials) increases. In general, it is considered good practice to try to graphically detect functional specification problems in econometric models preliminarily, even if it is also later verified in the econometric model (omitted for simplicity).

Figure 13. Quadratic adjustment between price and material costs



Source: prepared in-house.

393. In Figure 13, the solid line shows the predicted values of the price using exclusively the explanatory variable of material costs. The grey area represents the confidence intervals around the predicted value, indicating that there is a 95% probability that the population values lie within this range. Note that the reduced grey area indicated a high probability of a true quadratic relationship between these two variables. On the left are the relationships in Region A and B before and after the infringement, which indicates a similarity in the level and shape of the evolution of prices explained by the evolution of material costs in both regions during these periods. In contrast, during the infringement period, the price level in Region A is significantly higher than in Region B, despite the fact that the two markets show a similar relationship between prices and material costs.

394. Once the functional form has been specified, the quadratic relationship is introduced into the multiple regression model, which includes, in addition to the material cost variable, another variable, which is the squared cost of materials. The other variables are assumed to have a linear relationship with price. The equation [S_multiple] represents the abbreviated form of the estimated regression:

[S_multiple]

$$P_{ft} = \alpha + \beta \cdot cartel_f + \delta \cdot supply_{ft} + \gamma \cdot demand_{ft} + \varepsilon_{ft},$$

where supply and demand group the control variables shown in Section A4.2.3[221].

395. Table 8 shows that the estimated coefficient of the *cartel* variable is greater in the case of multiple estimation than in the simple one, while the corresponding standard error is smaller. This difference and the narrow range of the standard error suggests that the inclusion of more explanatory variables in the multiple regression model incorporates additional information that increases the precision of the estimation.

396. In addition, the results presented in the second column of Table 8 reveal that, as a result of including more relevant explanatory variables, the goodness-of-fit of the model improves substantially: the adjusted R2 of 0.58 obtained in [S_simple] rises to 0.77 in [S_multiple]. Note that we are referring here to adjusted R2, which is a preferable measure to R2, since when a new variable is added, the adjusted R2 value only increases if the variable increases the explanatory power of the model[222].

397. To check the **relevance of the set of explanatory variables included**, we perform a joint significance F-test the result of which indicates the relevance of the model[223]. We can reject the null hypothesis that the parameters of the explanatory variables included are jointly equal to zero, which implies that the set of explanatory variables included in the model is relevant for explaining the variable of interest (the price). It should also be noted that by introducing more explanatory variables, the standard error of the parameter β has been reduced in the [S_multiple] model with respect to the previous model, which indicates a greater efficiency in estimation. In principle, this model would be preferable to a simple regression, which provides little additional information compared to a simple comparison of means.

[221] The full form of the estimate is: $P_{ft} = \alpha + \beta \cdot cartel_f + \delta_1 \cdot C_{m_{ft}} + \delta_2 \cdot C^2_{m_{ft}} + \delta_3 \cdot P_{enel_t} + \delta_4 \cdot IND_{lab_{ft}} + \gamma_1 \cdot GDP_{rt} + \varepsilon_{ft}$

[222] For more details, see Subsection A2.2.4.2 on the goodness of fit.

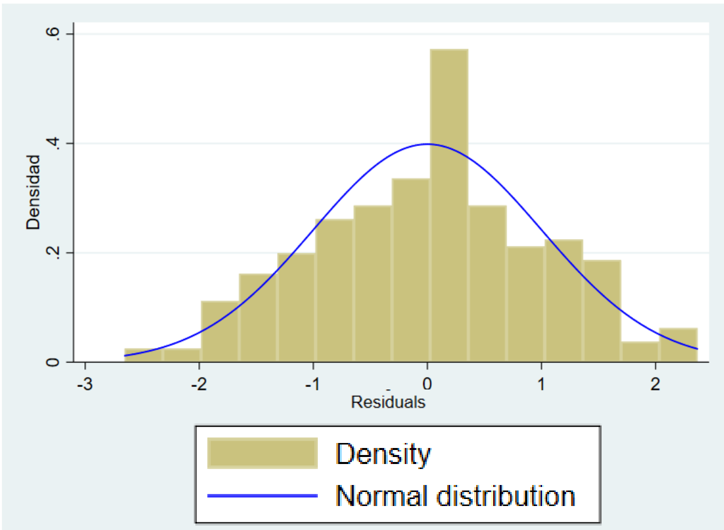
[223] $F(6; 233) = 845.89;\ Prob > F = 0.000.$

Checking the correct model specification

398. To assess whether the suppositions of the linear regression model are met, a series of checks are introduced that can be carried out with the results of the econometric estimation[224]. It should be noted that the checks introduced are not intended to be exhaustive and that, in reality, failure to comply with one or more of them does not necessarily invalidate the model.

399. First, the **distribution of the model residuals** [S_multiple] is analysed, namely, the difference between the actual values of the price and the values predicted by the model. In Figure 14 we can see, on the one hand, how the residuals are distributed (centred around zero and with a standard error slightly above one) in a histogram and, on the other hand, how the histogram can be compared to a normal distribution, and we can see that they both are close to each other[225].

Figure 14. Distribution of the residuals in the model [S_multiple]



Source: prepared in-house.
Note: The data used is for markets A and B between Jan. 2012 and Dec. 2013.

400. Furthermore, the study of the distribution of residuals and fitted values seems to fulfil the necessary criteria of randomness according to the graph, which indicates absence of heteroscedasticity (the variance of the sample errors seems to be constant). We can confirm this observation with the Breusch-Pagan test or with the White test in the case of a non-linear functional form, and the null hypothesis on the existence of

[224] See Sections 2.3 and 2.5 of Annex 2 to consult the suppositions and frequent problems, respectively, of the estimates by OLS.

[225] Subsection 1.2.1 of Annex 2 contains more information on the use of histograms.

homoscedasticity cannot be rejected. If we had evidence of heteroscedasticity, it would be common practice to use robust standard errors to obtain more accurate results[226].

401. Finally, the degree of **multicollinearity**[227] of the model, that is, whether and to what extent there is a linear relationship between the explanatory variables, is analysed **using the variance inflation factor** (VIF). If there is a strong linear relationship between the explanatory variables, their inclusion reduces the precision of the model and can alter the coefficients of interest, something that can be problematic, particularly if it affects the variable that captures the effect of the infringement. It can be seen that the GDP and cartel variables present higher VIF values (VIF_{cartel}=5.49; VIF_{GDP}=4.76), but these are below the threshold that is usually considered problematic[228].

402. After verifying the appropriate specification of the model, it can be concluded, with a confidence level of 99%, that the **average overcharge** associated with the cartel during the period of infringement in Region A equals 17.98 euros per unit of intermediate product sold. The explanatory variables that significantly impact the price level in these markets are the price of electricity, the costs of materials and the quantity of the intermediate product sold. Regarding the labour cost index, although its effect is not significantly different from zero, the variable is maintained in the model as it is considered a relevant factor when explaining the price of the product.

403. It should be noted that, when identifying the effects attributable to each explanatory variable, it is important to consider its functional relationship with the dependent variable (price). Electric r and labour costs have a linear relationship, so their (marginal[229]) effects on the price correspond to the estimated coefficients in Table 8 ("beta"). However, material costs affect prices quadratically, therefore, to determine their impact on prices, it is

[226] To expand on the consequences of heteroscedasticity and its possible solutions, see Subsection A2.2.5.4.

[227] In the case of perfect multicollinearity, all statistical programs automatically omit the variable in question, so the analysis should focus more on the study of imperfect multicollinearity.

[228] Although there is no consensus in the econometric literature, it is usually considered that there is a serious multicollinearity problem with VIFs greater than 10.

[229] Marginal effects tell us how the dependent variable changes when a certain explanatory variable changes (usually assuming all other explanatory variables remain constant).

necessary to simultaneously consider both variables corresponding to material costs (C_{mat} and C_{mat}^2)[230].

A4.4.1.3 Addressing the potential endogeneity problem: instrumental variable model.

404. **Endogeneity** is one of the most frequent problems in applied economics. As has been pointed out in Annex 2, endogeneity arises because we have elements included in the error term that are related to the explanatory variables included in the model.

405. To illustrate the endogeneity problem and a possible solution, **an endogenous control variable is added** to the multiple regression model (renamed [S_multiple*]): **the quantity of product** sold by each factory f in month t [Q_{ft}]. The inclusion of this variable would create a situation of bidirectional causality with the price of the product, since not only the variations in the quantity sold influence the price level, but also the price level determines the quantity produced[231]. This would violate one of the basic premises of OLS estimation (the non-correlation between the error and the explanatory variables), leading to inconsistent and biased estimates.

406. The endogeneity problem is **usually solved by including instrumental variables**, i.e., replacing the endogenous variable with a variable (or more than one) that must meet two criteria to be valid as an instrument: (i) It has to be **relevant**, that is, explain the price, and (ii) it cannot be correlated with the error term (it must be **exogenous**). In this case, the instrument used will be an *ad hoc* variable that will capture some **idiosyncratic characteristics of the factories** that can influence production at any given time [CAR_{ft}].

407. Before carrying out the estimation, it is necessary to verify that the instrument fulfils the two aforementioned conditions:

408. In terms of relevance, it is verified that there is a correlation between the instrument and the endogenous variable: the correlation coefficient between

[230] Specifically, its marginal effect is (approximately) equal to $\widehat{\delta_1} + 2\widehat{\delta_2} \cdot C_{mat_{ft}}$, i.e., the marginal effect changes depending on the cost of materials taken as a starting point. For example, if we start from the average cost of materials (16 euros), an increase of one euro in said cost would imply an increase of 0.63 euros in the price. On the other hand, if the cost increased from 17 to 18 euros, the price would increase by 0.7 euros, i.e., more than proportionally.

[231] The objective of this section is to show, in practical terms, the consequences of including endogenous variables in an econometric model and how this can be resolved. For this reason, it has been decided to introduce a clearly endogenous variable, although this does not mean that its inclusion in the hypothetical model presented is recommended. In fact, in a reduced form model, it would not be appropriate to include the price and the quantities sold at the same time.

the sales of the intermediate product and the characteristic variable of the factories is negative, $Corr(Q, CAR) = -0.6$.

409. The verification of the exogeneity condition is problematic as it cannot be verified empirically. Therefore, it is necessary to prove that the instrument is independent of the error term, based on knowledge of the market, the infringement and economic theory. It is assumed that the proposed instrument meets this condition, so that it is not related to unconsidered factors that affect the price of the intermediate product.

410. Once both conditions have been verified, a new model is estimated using the proposed instrument[232]. This is donde using Two-Stage Least Squares (2SLS) estimation, denoted as [S_instrument]. It is preferable to use 2SLS rather than OLS, because this method produces consistent estimates if the conditions described above (relevance and independence) are met[233]. This method is formalised as follows:

[S_instrument]

1st stage: $\widehat{Q_{ft}} = \alpha' + \beta' \cdot cartel_f + \delta' \cdot exogenous\ variables_{ft} + \gamma' \cdot instrument + \upsilon_{ft}$

2nd stage: $P_{ft} = \alpha + \beta \cdot cartel_f + \delta \cdot\ exogenous\ variables_{ft} + \gamma \cdot \widehat{Q_{ft}} + \varepsilon_{ft}$

411. In this case, two equations are used in the estimation. In the first stage, an auxiliary regression is estimated[234], in which the dependent variable is the endogenous variable, $[Q_{ft}]$, and the instrument and the rest of the explanatory variables of the original model are included among the independent variables (it is assumed that they are all exogenous). In the second stage, the original model is estimated, but substituting the values of the endogenous explanatory variable (sale of the intermediate product) by the values predicted in the auxiliary regression of the first stage. The estimation results are shown in the fourth column of Table 8[235].

412. A fact that could support the suspicion that the [S_multiple*] model has an endogeneity problem is that, by including a valid instrument (relevant and exogenous), the estimate of the coefficient of the potentially endogenous

---

232 Although in this example it may seem trivial, it must be considered that, in practice, it is not easy to find relevant and exogenous instruments and that the use of bad instruments can lead to a greater bias than that suffered when using OLS.

233 Mathematically, they are equivalent to: $Corr(Q; CAR) \neq 0\ y\ Corr(CAR; \varepsilon) = 0$

234 Stage 1 (auxiliary): $\widehat{Q_{ft}} = \alpha + \beta' \cdot cartel_f + \delta_1' \cdot C_{m_{ft}} + \delta_2' \cdot C^2_{m_{ft}} + \delta_3' \cdot P_{enel_t} + \delta_4' \cdot IND_{lab_{ft}} + \gamma_1' \cdot GDP_{rt} + \gamma_2' \cdot CARr_{ft} + \upsilon_{ft}$

235 Stage 2: $P_{ft} = \alpha + \beta \cdot cartel_f + \delta_1 \cdot C_{m_{ft}} + \delta_2 \cdot C^2_{m_{ft}} + \delta_3 \cdot P_{enel_t} + \delta_4 \cdot IND_{lab_{ft}} + \gamma_1 \cdot GDP_{rt} + \gamma_2 \cdot \widehat{\boldsymbol{Q}_{ft}} + \varepsilon_{ft}$

variable undergoes a significant change in relative terms, going from -1.15[236] to -2.06. However, to check whether the differences between the estimates [S_instrument] and [S_multiple*] are significant, it is common practice to apply the Durbin-Wu-Hausman test[237].

413. It is important to point out that **the standard error of an estimate with instrumental variables will always be larger than with OLS,** implying a larger confidence interval. That is, consistency is gained at the cost of reducing the efficiency of the model. Furthermore, the higher the correlation between the instrumented (endogenous) variable and the instrumental variables, the lower the standard error and the higher the estimation efficiency.

[236] To simplify the example, we have omitted the table showing the results of the estimation of the model with uncorrected endogeneity problems [S_multiple*].

[237] The Durbin-Wu-Hausman test evaluates the consistency of an estimator (S_multiple*) compared with an alternative estimator (S_instrument) that is less efficient but consistent. The result points to the lack of consistency of the model [S_multiple*] by rejecting the H0 that both coefficients are consistent (Chi2=16.22; Prob>Chi2=0.0002).

Table 8. Results of synchronous models

Dependent variable: price of intermediate product		S_simple	S_multiple	S_instrument
cartel	beta	17.363 ***	17.9766 ***	18.9247 ***
	standard error	0.7974	0.5519	0.6359
	t stat	21.7748	32.57	29.76
	p-value	0.0000	0.0000	0.0000
cost of materials	beta		-0.5201	-0.5842
	standard error		0.5823	0.5749
	t stat		-0.89	-1.02
	p-value		0.373	0.3096
costs of materials^2	beta		0.0349 ***	0.053 ***
	standard error		0.0154	1.0152
	t stat		2.26	3.49
	p-value		0.025	0.0000
wholesale price of electricity	beta		0.386 ***	0.318 ***
	standard error		0.0622	0.0621
	t stat		6.21	5.12
	p-value		0	0.0000
labour cost index	beta		0.039	0.023
	standard error		0.0657	0.0604
	t stat		0.59	0.38
	p-value		0.553	0.7038
GDP (regional)	beta		24.5744	18.9383 ***
	standard error		13.9182	14.7847
	t stat		1.77	1.28
	p-value		0.079	0.2002
constant	beta	27.105 ***	-9.9286	25.0553
	standard error	0.5638	11.2136	12.0488
	t stat	48.0729	-0.89	2.08
	p-value	0	0.377	0.0376
Number of observations		240	240	240
F test		474.14	846	
Prob>F		0.0000	0.0000	
R2		0.6658	0.7854	0.7521
R2_adj		0.6644	0.7741	0.7487

Legend: * level of confidence interval 90%; ** level of confidence interval 95%; *** level of confidence interval 99%

Source: prepared in-house.

A4.4.2 Analysis comparing different time periods (Diachronic analysis)

414. In the diachronic analysis, only the observations of the cartelised factories (region A) are applied; the evolution of their prices during the period of the cartel is compared with the periods not affected by the infringement. Below, two different approaches are presented to illustrate the diachronic analysis: the **dummy variable** approach and the **predictive** approach (including two

predictions of prices during the infringement: one from the earlier period and one from the later period).

A4.4.2.1 Dummy variable approach

415. At the centre of the analysis is the dummy variable $during_t$, which identifies the price premium in the cartelised period compared to the other unaffected periods: before and after the cartel. Thus, in this diachronic model, $during_t$ is equal to one when the infringement occurs, between January 2012 and December 2013, and equal to zero otherwise (January 2010 - December 2011 and January 2014 - December 2015)[238].

416. In the diachronic analysis, we propose two models with the same specification, the only difference being that, in the first, **[D_before]**, we compare the cartelised prices with those of the **pre-cartel period**, while in the second, **[D_after]**, they are compared with the prices of the **post cartel period**[239]. The regressions of the diachronic models with a dummy variable are formalised as follows:

[D_before] and [D_after]

$$P_{ft} = \alpha + \beta \cdot during_t + \delta \cdot supply_{ft} + \gamma \cdot demand_{ft} + \varepsilon_{ft}$$

417. In the following sections, for the sake of simplicity, detailed information on the checks conducted to ensure the fulfilment of OLS assumptions[240] will be excluded. Simultaneously, to facilitate the comparison across various models, the identical control variables from the [S_multiple] model outlined in Subsection A4.4.1.2 (adjusted for the specificities of each method) are incorporated.

[238] 240 observations are used: 48 months (24 from the cartel and 24 from the period before and after) multiplied by five companies (those from market A).

[239] In principle, there is nothing to prevent data from the pre-cartel and post-cartel periods from being used simultaneously (indeed, it would even be preferable if the data were available). However, it has been decided to show the results of models that use only one of the periods as a comparison scenario so that it is clear that, even in this simulated and ideal example, differences arise depending on whether data from before or after the infringement is used.

[240] It would be desirable for an expert report to analyse these issues.

Table 9. Results of the diachronic model estimations

Dependent variable: price of intermediate product	Diachronic dummy var.		Diachronic predictive			
	before vs during	during vs after	estimate of former period	prediction	estimate of former period	backcasting
	[D_before]	[D_after]	[D_predict]		[D_backcast]	
Predicted mean overprice (real price - prediction)				15.1293		14.9073
during	15.0436 ***	15.269 ***				
cost of materials	-0.3312	-0.6771	-0.342		-0.6514 **	
cost of materials^2	0.0571 ***	0.0684 ***	0.0551 ***		0.0644 ***	
wholesale price of electricity	0.3151 ***	0.4166 ***	0.0336 ***		0.0392 ***	
labour cost index	0.0354	0.0566	-0.086		0.0201	
GDP (regional)	11.4668	15.2479				
constant	-3.4961	0.0169	5.5262		-3.9877	
Number of observations	240	240	120		120	
F test	986.33	1535.73	186.57		339.23	
Prob>F	0.0000	0.0000	0.0000		0.0000	
R2	0.7761	0.7821	0.7303		0.7322	
R2_adj	0.7502	0.7789	0.7245		0.7300	

Legend: * level of confidence interval 90%; ** level of confidence interval 95%; *** level of confidence interval 99%

Source: prepared in-house.

418. The results corresponding to the dummy variable approach (presented in the second and third columns of Table 9) reveal that, in market A, if the influence of other determinants of the product price is kept constant, **the prices during the cartel period were 15.04 euros and 15.27 euros higher than in the periods without infringement** (before and after, respectively). It is also found that the variables are jointly significant and explain a high percentage of the price variation, with an adjusted R2 always above 75%.

A4.4.2.2 Predictive approach

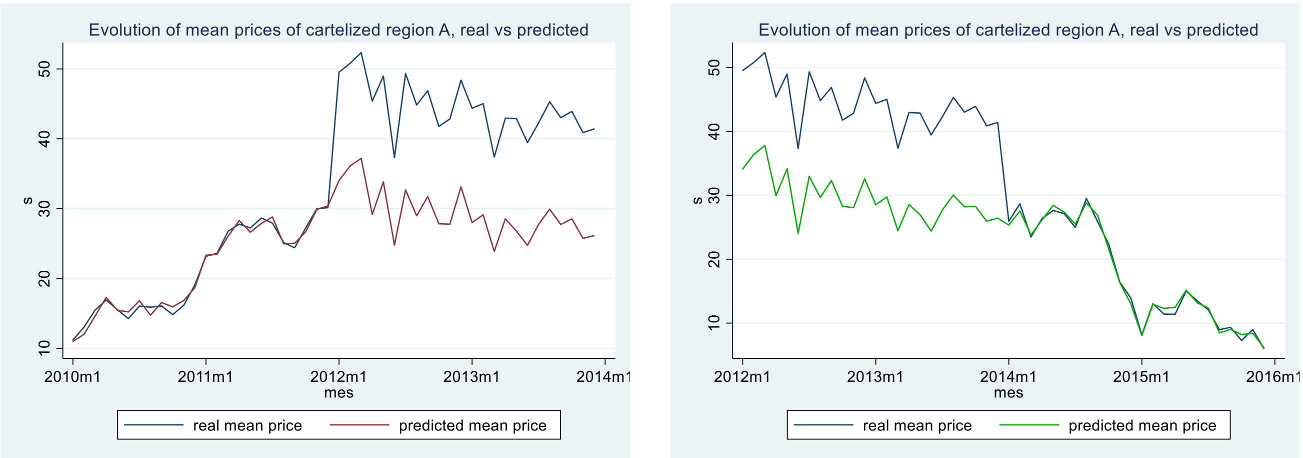
419. The predictive approach presented here comprises two estimates based on similar reasoning. The first is based on **pre-cartel** observations to project a hypothetical price development during the infringement to find the difference between the infringement and the actual evolution observed (forecasting). The second does the same using **post-cartel** observations (back-casting). The logic is that a model that includes all relevant explanatory variables and predicts prices below those actually observed in the absence of infringement points, would reveal the existence of overcharging.

420. To estimate the hypothetical evolution of prices in the absence of the infringement, the same specification of the previous model with a dummy variable is used. However, in this case the estimates exclusively use data

from the period before or after the infringement[241]. Once the parameter estimates are obtained, the hypothetical (predicted) prices are projected using the data of the explanatory variables in the cartelised period. The overcharge is calculated as the difference between the actual and predicted prices. The last four columns of Table 9 show the results.

421. The estimates give results with a goodness-of-fit greater than 70% and indicate that the overcharge during the cartel was 15.13 euros and 14.91 euros higher than in the previous and subsequent periods, respectively. The following two figures illustrate the difference between the hypothetical mean evolution and the real prices in market A.

Figure 15. Linear price prediction from data before (left) and after (right) the cartel



Source: prepared in-house.

A4.4.3 Difference-in-differences analysis

422. The difference-in-differences (DID) method combines elements of synchronous and diachronic analysis. Firstly, the evolution of factory prices in periods not affected by the infringement is compared in both regions to obtain the temporal difference $[Dif_t]$. Then the prices of the cartelised and non-cartelised factories are compared to discover the difference between the groups of factories in both regions $[Dif_r]$. Finally, the second level of differences is calculated, subtracting the temporal difference and the difference between regions $[DID = Dif_r - Dif_t]$.

423. An important advantage of using the DID method is that it allows comparisons to be made between the different markets even when the

241 In this approach,120 observations are used: 24 months (those of the period before and after the infringement) multiplied by five companies (region A).

number of control variables is reduced, as long as it is reasonable to assume that the unobserved differences between the two markets ("unobservable heterogeneity") remain constant over time[242]. In this way, it is assumed that the evolution of the unobservable control variables that are not explicitly included in a DID estimate is similar both in the region where the infringement took place and in the comparators.

424. To capture possible influences of specific factors not observed at the market level or from a temporal perspective, dummy variables are included for each region and month ("fixed effects"). The inclusion of these two fixed effects is a commonly used extension of the DID method, which is known as *two-way fixed effects*)[243].

A4.4.3.1 Preconditions

425. Once the methodological framework has been defined, to apply the DID methodology it is necessary to satisfy the **assumption of parallel trends** between the units that are affected by the infringement and those that are not affected by the infringement. If this condition, which must be reasonably justified based on the particular circumstances of the case, is met, it can be argued that, in the absence of the infringement, the evolution of the variable of interest (in this case, prices) of the units affected would have been the same as that of unaffected units. This allows us to project a hypothetical evolution of the "affected" observations during the infringement and calculate the difference between this projection and the actual observations.

426. In our case, as an illustration, Figure 11 shows for illustrative purposes we present the evolution of the average prices of the factories in the two markets in Figure 11, where there is clear deviation from the parallel trend only during the existence of the cartel. In addition, the change in the evolution of prices in market A occurs suddenly at the beginning of the cartel and disappears when it ends.

427. In order to have not only visual but also quantitative evidence, a test is performed to **assess the existence of parallel trends in the periods before and after the infringement**. This is done by running a regression

[242] For more details, see Subsection 2.3.1.c of the Guide.

[243] This extension of the DID method has the advantage of enabling a consistent estimation of mean treatment effects incorporating multiple periods, variation over time of the treatments and the fixed effects according to units. In other words, it allows considering heterogeneous effects between units and over time through fixed effects. However, it also requires a large number of observations for its execution as a consequence of the number of fixed effect variables created.

on the data from all periods, where the dependent variable is the price and we include three types of binary explanatory variables:

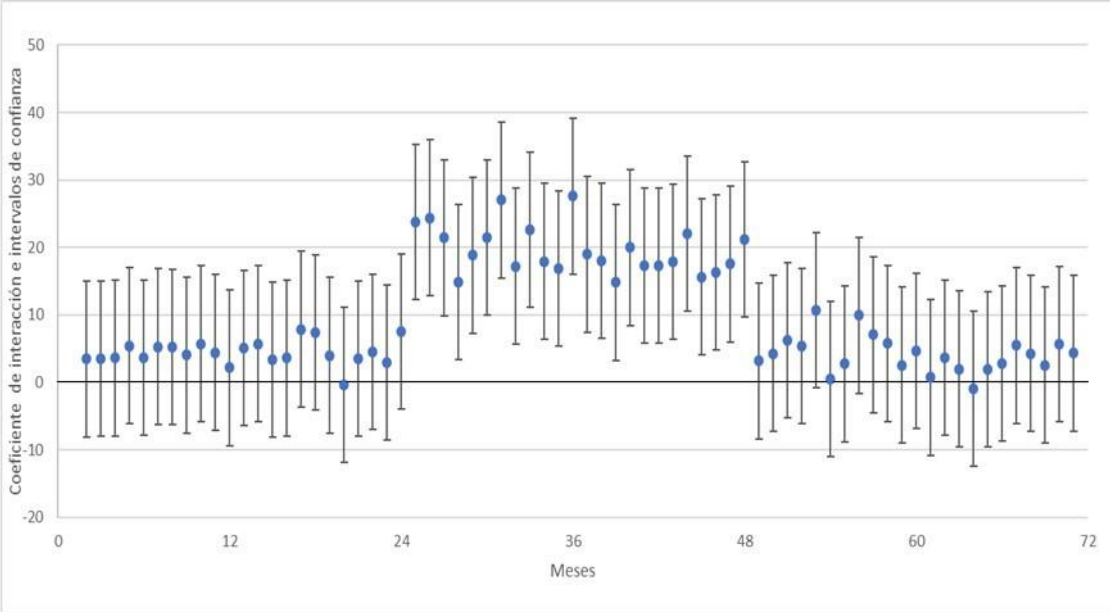
i. the **time-fixed effects** (one variable for each month), which are intended to capture price differences between different periods common to the regions $[\eta_t]$,

ii. the **fixed effects of the infringement** (a variable that indicates whether the observation belongs to cartelised region A, or to non-cartelised region B), which are intended to capture price differences between the units affected and not affected by the infringement, invariant over time $[\lambda_r]$, and

iii. the **interaction between the time-fixed effects and of the fixed effects of the infringement** (that is, there is a variable for each month that is the result of multiplying η_t and λ_r). The coefficient of this interaction will be **relevant for the purposes of testing the assumption of parallel trends**, after taking into account the rest of the fixed effects.

428. When estimating the model, we find that the estimated values of the interaction terms are not significantly different from zero in the pre- and post-cartel periods. This suggests that, both before and after the infringement, factory prices in the two markets behaved similarly, once regional, and monthly differences were controlled. This result can be better understood from the following graph, which shows that, at the 1% level of significance, the estimates for all months before and after the infringement (blue dots) contain the value zero within their confidence intervals (vertical bars), that is, it cannot be denied that the prices in both markets were similar (or, in other words, that the difference between both prices was nil) in the period not affected by the infringement. It is also observed that the price difference during all the months of the infringement is positive (the confidence intervals do not contain the value zero), suggesting that preliminary evidence is favourable to the existence of overcharge.

Figure 16. Testing the parallel trends assumption



Source: prepared in-house.

A4.4.3.2 DID model estimation

429. Once the probable fulfilment of the parallel trends' assumption has been justified, the DID method is applied by introducing a dummy variable $[\boldsymbol{DID_{ft}}]$ that takes the value of one if an observation corresponds to the cartelised factories during the months of the infringement, while, in any other case, the value of the variable is zero. This is the result of the interaction (multiplication) of the dummy variables $\boldsymbol{cartel_f}$[244] and $\boldsymbol{during_t}$[245]. Thus, for the purpose of quantifying the damage, the parameter of interest $\boldsymbol{\delta}$ is found through the interaction of these variables. This parameter identifies the average difference between the prices of the cartelised market and non-cartelised market before and during the cartel. The same model is also estimated using data only from during and after the infringement.

$$\textbf{[DID]}\ P_{ft} = \alpha + \beta \cdot during_t + \gamma \cdot cartel_f + \delta \cdot DID_{ft} + \rho \cdot X_{ft} + \lambda_r + \eta_t + \varepsilon_{ft}$$

430. The term X_{ft} includes several explanatory variables (electricity, material costs, labour cost index and regional GDP). The constant is represented by $[\alpha]$, while the region and time fixed effects mentioned above, are $[\lambda_r]$ and

244 It takes a value of one if the factory belongs to the cartelised region (A) and zero if it belongs to another region (B).

245 It takes values of one if the observations belong to the period during which the cartel was active and zero outside those periods.

[η_t], respectively. The error term is [ε_{ft}]. The results are shown in the table below.

Table 10. Results related to the DID model.

Dependent variable: price of intermediate product	[DID_1]		[DID_2]	
DID	15.6893	***	15.0143	***
cost of materials	-0.4275	*	-0.4662	**
cost of materials^2	0.0579	***	0.0569	***
wholesale price of electricity	0.2949	***	0.3772	***
labour index	0.0825		0.0740	
GDP	16.9458		27.5843	
constant	-5.5435		11.1177	
Fixed effects: time	yes		yes	
Fixed effects: markets	yes		yes	
Number of observations	480		480	
F test	408.45		567.51	
Prob>F	0.0000		0.0000	
R2	0.8325		0.8362	
R2_adj	0.8269		0.8302	

Legend: * level of confidence interval 90%; ** level of confidence interval 95%; *** level of confidence interval 99%

Source: prepared in-house.

431. The results of the two DID estimates reveal that during the infringement, prices were around 15 euros higher in market A than in market B. The explanatory variables used present a significance level of 99% and the **F statistic** suggest that the fixed time and market effects[246] are significantly different from zero[247].

Summary of the estimation results

432. In short, throughout this section, the average overcharge attributable to the cartel in market A has been calculated using different methods (synchronous, diachronous with a dummy variable approach, predictive, and difference-in-differences) that lead to differences in the results that are

[246] In this example, the fixed effects of markets and regions coincide, since all the factories in Region A participate in the cartel and all the factories in Region B do not. In general, in DID models, the fixed effects are used to control for the differences (heterogeneity) not observed at the temporal level or in the units analysed.

[247] We reject the H0 that the estimated parameters of the fixed effects are equal to zero.

in a range of less than four euros, as can be seen in Table 11. Within the framework of this example with dummy data, the highest price overcharge (17.98 euros/unit) is identified using the synchronous method that compares the evolution of the prices of cartelised factories with non-cartelised ones during the infringement, while the lowest overcharge is obtained with forward diachronic prediction (14.91 euros/unit)[248].

Table 11. Summary of the results for the different methods

Method	comparative periods	euros/unit
Synchronous (multiple regression model)	during	17.98
Diachronic: dummy variable approaches	before vs during	15.04
	during vs after	15.27
Diachronic: predictive approaches	from previous period forward	15.13
	from later period backwards	14.91
Difference-in-differences	before vs during	15.69
	during vs after	15.01

Source: prepared in-house.
Note: the overcharge is rounded to two decimal places.

A4.5 CAPITALISATION OF DAMAGE

433. As indicated throughout the Guide, once the overcharge has been calculated, it is necessary to **express it in current value using a capitalisation rate**.

434. The estimated values in the previous sections represent the annual overcharge suffered by the buyers of the intermediate product. Therefore, when calculating actual loss suffered, the amount that the plaintiff acquired from the intermediate product each year is taken into account. Let us suppose, as indicated in the following table, that in the first year 100,000 units of the cartelised product were purchased and in the second year, 75,000 units were bought, and that the average price overcharge incurred was 15.69 euros per unit[249]. The actual loss corresponding to each of these years is therefore obtained by multiplying the quantities purchased by the

248 Although the overcharge is expressed in euros per unit, it is common for it to appear as a percentage of the actual or counterfactual price. Sometimes the variables are expressed in logarithms to facilitate their interpretation in that sense (see Section A2.2.1 with examples of interpretation of the coefficients according to the functional form).

249 As an example, we take the result of the estimation of difference-in-differences, before/during.

quantified overcharge. To simplify the analysis, it is assumed that the overcharge was paid on January 1 of each year, so that only annual interest needs to be calculated[250].

435. Once the harm for each period has been quantified and expressed in monetary units, it should be converted into a single amount claimed, in other words, it has to be capitalised by applying the capitalisation rate considered appropriate. In this example, it is assumed that the plaintiff chooses to compound the harm until the claim is filed using the weighted average cost of capital (WACC), which is assumed to be constant and equal to 3% per year[251].

Table 12. Harm capitalisation using the DID_ before / during model as an example.

Cálculo	Conceptos	medidas	**Capitalización compuesta**	**capitalización simple**	**capitalización compuesta (menor tasa de capit.)**
A	Sobreprecio estimado	€/unidad	15,6893	15,6893	15,6893
B	Cantidad del producto comprado en 1er año	unidades	100.000	100.000	100.000
C	Cantidad del producto comprado en 2º año	unidades	75.000	75.000	75.000
D	Daño emergente en términos nominales	€	2.745.628	2.745.628	2.745.628
E	Tipo de interés (igual para todos los años)	%	3,0%	3,0%	**1,5%**
F=A*B*E	Capitalización tras el 1er año de infracción	€	47.068	47.068	23.534
G=D+F	**Valor inicial del daño al terminar la infracción**	**€**	**2.792.695**	**2.792.695**	**2.769.161**
H	Nº años transcurridos desde la finalización de la infracción y la presentación de la demanda	años	3	3	3
I=G*E	Capitalización del 1er año	€	83.781	83.781	41.537
J=(G+I)*E	Capitalización del 2º año	€	86.294	83.781	42.160
K=(G+I+J)*E	Capitalización del 3er año	€	88.883	83.781	42.793
L=I+J+K	**Total capitalización del período A (INTERESES)**	**€**	**258.958**	**251.343**	**126.491**
M=G+L	**Valor final (DAÑO EMERGENTE + INTERESES del período A)**	**€**	**3.051.654**	**3.044.038**	**2.895.652**
N	**Tipo de interés**	**%**	3,0%	3,0%	1,5%
O=M*N	**Capitalización del período B**	**€**	91.550	83.781	43.435
P=M+O	**DAÑO EMERGENTE + INTERESES**	**€**	**3.143.203**	**3.127.819**	**2.939.087**

Source: prepared in-house.

436. In the first stage, the actual loss for the first year of the cartel is capitalised, resulting in an initial value of the harm at the end of the infringement of 2,792,695 euros, expressed in euros for the year in which the infringement came to an end.

[250] In practice, it is necessary to take into account the specific time at which the harm occurred, since this will affect the interest calculation (logically, two amounts from the same year cannot be capitalised in the same way if one corresponds to the 1st of January and another to 31st December, especially if the capitalisation rate is high).

[251] It is desirable that the expert reports justify both the interest rate applied and the way it was calculated. For more information, see Section 2.4 of the Guide.

437. In the second stage, it is necessary to capitalise the amount up to the date of the claim is filed (period A). It is assumed that, after the end of the infringement, 3 years pass until the injured party files their claim. Therefore, the compound capitalisation of the harm is calculated for each of these years, adding each year the interest accrued in the previous periods. In this way, it is calculated that the interest amounts to 258,958 euros, as illustrated in Table 12 (row L). In this way, the sum of the actual losses and interest corresponding to period A reaches 3,051,654 million euros (row M) (the amount claimed at the time of filing the claim).

438. Thirdly, to obtain full compensation for the damage, it is also necessary to consider the interest corresponding to the period between the filing of the claim and the ruling of the first instance (period B). Assuming that the same interest rate (3%) is applied to compensate for the time elapsed during the legal proceedings (assumed to be one year), a total harm (actual loss and interest) of 3,143,203 euros is obtained, as can be seen in Table 12 (row P)[252].

439. If there is a delay in the payment of the compensation from the ruling of the first instance, the interest rate related to the procedural delay should be applied, although, as stated in Subsection 2.4.3 of the Guide, it is applied *ex officio* by the court, without the need to request it in the legal action.

440. Finally, the possible impact on the harm calculation of modifying two fundamental aspects of capitalisation **-the calculation method and the magnitude of the interest rate-** is analysed[253].

1. First, following the example of the DID before/during model (DID1), **it is possible to propose that the capitalisation be calculated in a simple way**, in such a way that the interest generated in each period is applied only to the initial capital, without considering the interest of previous periods. This would result in a total interest of 335,124 euros, which is 4.4% less than in the compound capitalisation scenario (350,508 euros). The difference between the two capitalisation methods is greater the longer the period between the onset of the harm and the ruling of the first instance.

2. Secondly, it can be assumed that the capitalisation method is still compound, but that **the rate used to calculate the interest rate for periods A and B is lower**: it is no longer 3% but, for example, 1.5%.

[252] The interest rate for period A has also been taken into account in the calculation, as it involves compound capitalisation.

[253] The details are in the last two columns of Table 12.

Under this assumption, the calculation method would be the same, but the total interest would be 169,926 euros, 51.5% lower than in the baseline scenario. This illustrates the importance of applying an appropriate and justified rate in order to obtain full compensation for the damage.

A4.6 CONCLUSIONS FROM THE PRACTICAL EXAMPLE

441. The objective of this practical exercise was to highlight, in a non-exhaustive manner, certain necessary precautions, in order to guide the preparation and evaluation of an expert report on the quantification of harm. The methods presented here focus on techniques based on the fields of statistics and econometrics, to familiarise the reader with the way of arguing and reasoning in this field. However, the example has been constructed in a simplified way to illustrate various methods and techniques explained during the Guide, so the data used has been artificially generated for this purpose.

442. The different **methods** shown can be **complementary** and share common assumptions, but each one has its framework of analysis, which **does not necessarily lead to the same result**. It is more common for the results of the different methods to differ from each other, which allows us to determine a range (interval) of possible effects. In the example, the different methods result in an overcharge of between 14.91 euros/unit and 17.98 euros/unit.

443. To select the most reliable results, an analysis should be carried out taking into account aspects such as **the choice of the counterfactual, the delimitation of the period, the handling of the database or the pass-on of costs, among other things**. In the present case, given that this is an example of artificially generated data and that, in addition, a number of simplified assumptions have been made, it is not possible to carry out a comparative analysis of the various overcharges calculated.

444. As indicated in this Guide, in cases where the results of the different approaches deviate widely and the different assumptions applied in each method make comparison difficult, **the reasons for the differences found** should be highlighted. Likewise, one should also question whether the results obtained constitute a **minimum or maximum value** of the harm caused by the infringement.

445. Finally, the example also shows the importance of correctly capitalising the harm to arrive at full compensation (it has been shown how, depending on the capitalisation rate and calculation method chosen, the final result of the compensation can differ considerably).