A critical examination of collusion, the pricing behaviour of a multi-product cartel and the cartel enforcement record in South Africa

by

Tapera Gilbert Muzata

Dissertation presented for the degree of Doctor of Philosophy (Economics)
at Stellenbosch University

Promoter: Prof. Willem Boshoff

Faculty of Economic and Management Sciences

Department of Economics

March 2021

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication by Stellenbosch University will not infringe on any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

This dissertation includes 0 original papers published in peer-reviewed journals or books and 3 unpublished publications. The development and writing of the papers (published and unpublished) were the principal responsibility of myself and, for each of the cases where this is not the case, a declaration is included in the dissertation indicating the nature and extent of the contributions of co-authors.

March 2021

i

Abstract

In this dissertation three research questions relating to collusion and cartel enforcement in South Africa were examined. The first question entailed examining the characteristics of detected cartels, together with the institutional features of selected key South African cartels. The author found that South African cartels incorporated some of the institutional features reported in the literature, including compensation schemes, joint ventures, sub-contracting arrangements, and entry or expansion deterrence strategies. In some selected key cartels, firms participated in collusion at two successive levels of the value chain, giving the cartels greater control over pricing throughout the value chain. Communication and monitoring were found to often involve a mix of various forms, including firm-level mechanisms (notably sales infrastructure) that complemented the centralised communications mechanisms discussed in the literature. To help fully explain collusion under conditions of imperfect information, the theory should account for the complementarities among various forms of communication and monitoring used by cartels. In addition, the author found trade policy to be an inexpensive tool used by some cartels to weaken threats from imports.

The second question concerned the pricing dynamics of a cartel involving multi-product firms and where the cartel faces periods of instability, producing distinct collusive phases. Like single product cartels, a multi-product cartel raises prices above competitive levels, but to varying degrees on different products. Cartel overcharges also vary over collusive phases, influenced by the demand and supply conditions in each phase. This suggests that a multi-product cartel maximises profits by imposing overcharges that vary by product and over collusive phases in response to changing market conditions. The assumptions about the nature of the transition between collusion and competition affects overcharge estimates. This dissertation provides arguments that penalties and damages estimates, reflecting overcharges, should consider product-level and phase-specific overcharges, rather than relying on averages.

Finally, the dissertation examined the cartel enforcement record from a deterrence perspective, focusing on the drivers of cartel enforcement, the duration of cases from initiation to final decision, and the subsequent impact on the deterrence-effect of penalties. Leniency, settlements and penalties, supported by increased funding for the Competition Commission are considered to be the main drivers of cartel enforcement in South Africa. Contrary to expectation, the author found that these have not reduced the duration of cases. Instead, case duration increased progressively over the study period. Delays in penalising firms have resulted in firms paying

significantly discounted penalties, weakening the deterrence effect of penalties. To preserve the deterrence effect of penalties, the author argues that an optimal cartel enforcement policy should account for these delays and should focus on higher present-value penalties for those firms that delay finalising cases.

This dissertation, focusing on South Africa, contributes to the body of empirical literature on collusion and cartel enforcement. It provides suggestions for further advances in the theory on collusion under imperfect information and on overcharge estimation when dealing with multi-product and multi-period collusion. The dissertation also makes policy contributions that could enhance the efficacy of cartel enforcement in South Africa and in other jurisdictions.

Opsomming

Hierdie proefskrif ondersoek drie vrae wat met samespanning en die toepassing van kartelwetgewinig in Suid-Afrika verband hou. Die eerste vraag behels die eienskappe van kartelle wat reeds ontbloot is, tesame met die institusionele kenmerke van uitgesoekte kartelle. Die proefskrif bevind dat Suid-Afrikaanse kartelle sommige van die institusionele eienskappe in die literatuur weerspieël, insluitend vergoedingskemas, gesamentlike ondernemings, subkontrakteringsooreenkomste, en strategieë om toetrede of uitbreiding te voorkom. In sommige sleutelgevalle word bevind dat firmas op twee opeenvolgende vlakke van 'n waardeketting saamspan, ten einde beter beheer oor prys te bekom. Die proefskrif bevind dat kommunikasie en monitering dikwels 'n kombinasie van vorme behels, insluitend firma-vlak meganismes (insluitend verkoopsinfrastruktuur) wat gesentraliseerde kommunikasie-meganismes, soos in die literatuur vervat, komplementeer. Die proefskrif bevind dus dat teorie samespanning onder toestande van onvolledige inligting beter kan beskryf indien dit vir hierdie komplementariteite voorsiening maak. Voorts bevind die proefskrif dat sommige kartelle handelsbeleid as 'n goedkoop afweermetode gebruik om die bedreiging van invoere die hoof te bied.

Die tweede vraag hou met die prysdinamika van 'n kartel verband, waar multi-produk-firmas saamspan en waar die samespanning periodies onderbreek word. Soos vir samespanninig rondom 'n enkele produk, bevind die proefskrif dat 'n multi-produk-kartel pryse bo die mededingende vlak verhoog, maar in wisselende mate na gelang van die bepaalde produk. Voorts bevind die proefskrif dat pryse oor verskillende fases van samespanning wissel, na gelang van veranderende markomstandighede. Aannames omtrent die aard van die oorgang tussen fases van samespanning en van mededinging, en die impak daarvan op skattings van kartelskade, word ook ondersoek. Die proefskrif betoog dus vir kartelboetes en skadebepaling om op produk-vlak en fase-spesifieke-ontledings, eerder as gemiddeldes, staat te maak.

Laastens ondersoek die proefskrif die karteltoepassingsrekord vanuit 'n ontmoedigingsoogpunt. Die soeklig val op die drie plekke, naamlik die vernaamste drywers van karteltoepassing, die duur van kartelsake vanaf inisiasie tot finale besluit, en die gevolglike impak van hierdie tydsduur op kartelboetes. Aansoeke om toegeeflikheid, skikkingsooreenkomste en boetes – sowel as beter befondsing – word as die vernaamste drywers van karteltoepassing beskou. Teen verwagting bevind die proefskrif dat hierdie drywers nie die duur van kartelsake verminder het nie. Trouens, die duur van kartelsake, oor

die studieperiode, het stelselmatig toegeneem. Die proefskrif bevind voorts dat vertragings in die penalisering van firmas die voorkomende effek van boetes verlaag. Ten einde die impak van boetes te verbeter behoort karteltoepassing rekening met die kontantwaarde van boetes te hou, veral waar firmas die finalisering van kartelsake vertraag.

Hierdie proefskrif, met 'n fokus op Suid-Afrika, lewer dus bydraes tot die empiriese literatuur rondom samespanning en karteltoepassing. Die proefskrif doen voorstelle aan die hand vir verdere ontwikkeling van die teorie van samespanning onder onvolledige inligting en die skatting van kartelskade waar dit multi-produk- en multi-periode-samespanning aangaan. Die proefskrif doen ook voorstelle aan die hand om die doeltreffendheid van kartelbeleid in Suid-Afrika en in ander jurisdiksies te verhoog.

v

Acknowledgements

I have a lot of gratitude to God for making it possible to undertake these studies and for giving me the financial resources and strength to persevere in circumstances where I have had to balance a wide array of demanding tasks and during difficult times.

I have profound gratitude to Prof. Willem Boshoff for his patience, intellectual support, and comments on the various pieces of work that eventually became part of this dissertation. Any shortcomings are not of his making, but are mine.

Making progress entails facing several obstacles, chief of which are a busy schedule and responsibilities. This requires immense sacrifice and all the support one can get. The choices and trade-offs involved are not easy and have real consequences. In this regard, I am eternally grateful for all the support and encouragement I received from various people and organisations.

The MultiChoice Group paid my tuition fees, and I am indebted to the Group CEO, Mr Mawela, and the Chairman of the Board, Mr Patel, for enabling this to happen. All the odds were against it, but they supported me and made it happen. Ms Ndlovu afforded me some flexibility in balancing work and studies. I am thankful to her for her feedback, including for the comments that I received from practitioners, who participated in the ESSA conferences held in 2017 and in 2019.

The support and sacrifices of my family deserve mention. They have put up with hours, days, and months of my absence, as I worked on this dissertation. My wife has been incredibly supportive in every respect, including carrying a significant amount of the workload in home-schooling our children, during a crucial time in my studies. Ethan (7) and Theo (4) have been excellent in affording me much-needed quiet and privacy when I needed to concentrate. My brother more than did his part, from conceptual advice to pushing me to complete my studies and encouraging me at every opportunity.

Table of Contents

List of Figures

List of Tables

Chapter I

1.1 Introduction

Collusion has been a pervasive phenomenon across most sectors of the South African economy. In some sectors, such as the construction sector, pervasive collusion pushed the Competition Commission (Commission) to develop a sector-wide mechanism to resolve the cases that it was investigating. Figure 1.1 shows the distribution of the Commission's cartel enforcement activities across various sectors, based on the number of firms prosecuted for cartel conduct between April 1999 and March 2018.

Figure 1.1: Distribution of cartel prosecutions across sectors
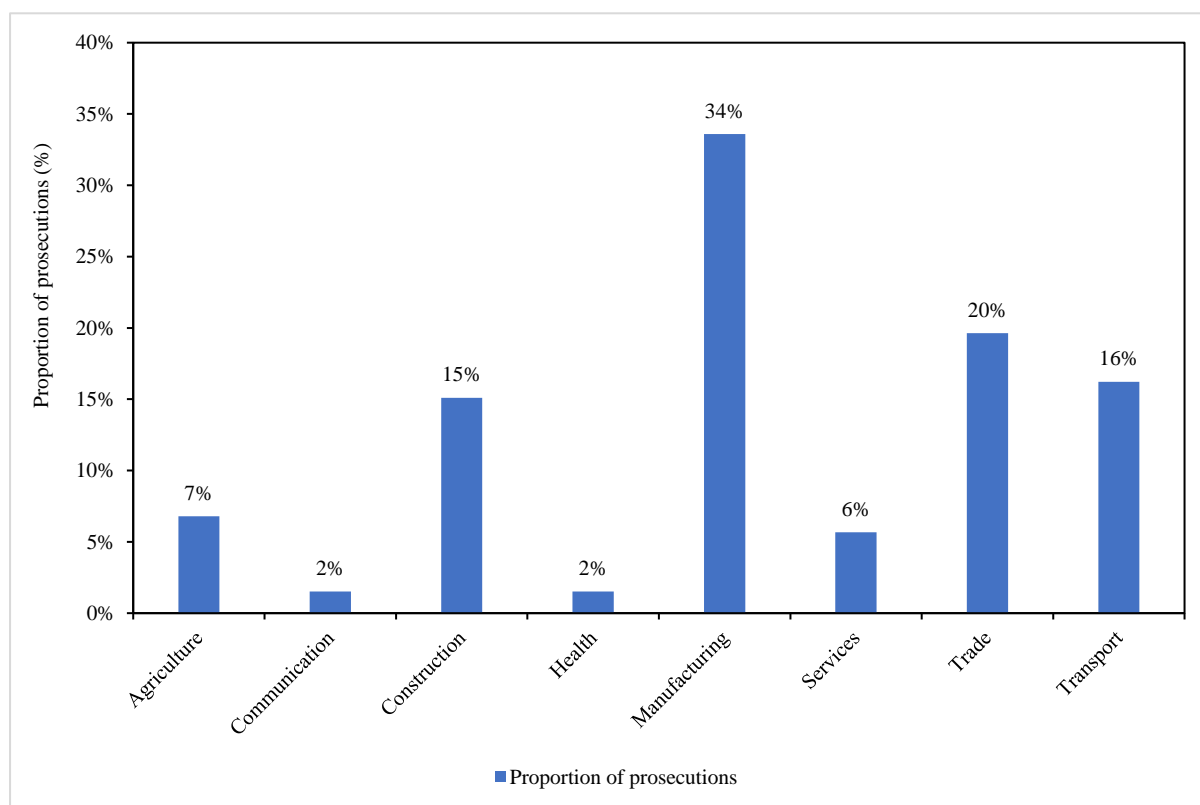
Source: Prosecutions data used in Chapter 4.

A significant number of cartels involved collusion across multiple products or in multiple markets, in some cases cartelising successive parts of the relevant value chains.[1] For example, the World Bank (2016) reported that for cartels prosecuted between 2005 and 2015, 63 per cent of the cases involved a firm that had also participated in another cartel in a different market of the same sector or in another sector. In Chapter 2, it is estimated that at least 42 per cent of

---

[1] See for example Tribunal case numbers 017277; 016972; 10/CR/Mar10; 15/CR/Mar10; and 15/CR/Feb07.

firms prosecuted between 1999 and 2018 either colluded more than once or belonged to a group in which another company was prosecuted for collusion in another market. Network mapping used by the World Bank (2016) shows extensive links between firms that partook in collusion in the agriculture and food sectors. This suggests that collusion could be a viable rational business strategy for multi-product/-market firms.

Multi-product firms that take part in collusion can cause (a) wide-ranging harm to consumers, and (b) medium to long-term damage to competition across markets. It may take a long time for the effects of entry barriers enacted by the cartels to dissipate, and to fully restore market dynamism. Given the multi-faceted nature of collusion involving multi-product firms, the nature and types of consumers that are harmed by the conduct will also likely vary significantly. Deterring multi-product collusion requires consideration of the extent of harm caused by the cartels, when imposing penalties. This applies also to any follow-on damages that consumers may pursue against the cartel members. Both deterrence and damages require a proper understanding of the pricing dynamics of a multi-product cartel. Most of the existing studies on cartel harm focus on individual product markets and do not examine collusion by multi-product firms as a portfolio of cartels [see, for example, Kovacic, Marshall, and Meurer (2018)].

Because of the pervasive nature of collusion in South Africa, the Commission has prioritised its resources to enhance its ability to detect and prosecute firms for collusion. This included introducing a leniency policy coupled with a settlement procedure, developing a prioritisation framework, setting up a standalone Cartels Division, increasing penalties over time, and using remedies designed to trigger competition in some of the markets where collusion was detected. Evidence supporting the use of active enforcement to end operating cartels includes Levenstein and Suslow (2011), who found that increased cartel enforcement aimed at international cartels raised the probability of ending cartels. This expansion was enabled by dedicating increased resources to competition authorities and by employing policy tools aimed at detecting and prosecuting cartels (e.g. leniency and settlement).

Leniency enhances cartel detection and prosecution by placing firms in asymmetric positions through offering avoidance of penalties in exchange for information on cartel conduct [Motta (2008)]. Firms are expected to have the incentive to settle cases because it results in lower penalties and saves them resources; although, as argued in this study, incentives relating to the timing of settlements may be ambiguous. Firms that settle with a competition authority

typically offer to co-operate by supplying information on the conduct of the cartel(s), as well as by supplying factual witnesses to enable the competition authority to prosecute fellow cartel members. Together, leniency and a settlement procedure are expected to be attractive to both firms and the competition authority [Edwards and Padilla (2010)].

Leniency and settlement procedures enable firms either to avoid sanction or to face reduced penalties, while the competition authority is enabled to detect cartels and to save time and financial resources, which it can then divert to detecting other cartels. Time and resource savings are expected to flow from the speedier and more effective resolution of cases, as the competition authority can gather more complete evidence on the conduct involved. This, in turn, is expected to result in the remaining cartel members being more willing to settle cases or to face a higher probability of prosecution and penalties. The general expectation is that all these enforcement initiatives increase the probability that collusion will be detected and prosecuted, and that deterrence will be achieved. As this study shows, leniency and settlements have played a major role in cartel enforcement in South Africa.

Recent evidence suggests that there is a decline in leniency applications, meaning that the future of effective cartel enforcement requires innovations in cartel detection. One area of innovation in cartel detection relates to cartel organisation and management. Cartels, by nature, are secretive. Understanding the characteristics of collusion in South Africa and its institutional structure can help competition authorities better identify patterns of firm and industry behaviour that are likely to be more consistent with collusion than with the independent actions of competing firms. Prosecution decisions and case studies of selected South African cartels provide useful information on the organisation and management structure of the cartels. This information includes the nature and characteristics of the cartel's conduct, communication within the cartels, information revealing how firms may have maintained incentives for collusion, monitored and deterred unilateral defection by cartel members (commonly called cheating), and restricted external threats. Examining the characteristics of collusion in South Africa and the institutional features of key selected cartels can complement behavioural and structural screens in detecting collusion.

1.2 Problem statement, research goals and questions

The South African experience with cartels shows that they were widespread across the various sectors of the economy, in many instances lasting several years, and empirical evidence from

studies of specific cases suggests that cartels exercised cartel market power. This experience is not unique to South Africa. As a result, theoretically, empirically, and from a policy perspective, debates have rightly moved on from those in the earlier years of Stigler's (1964) suggestions that collusion is inherently unstable. It has become widely accepted that cartels can be stable, and firms have powerful incentives to participate in collusion [e.g. Levenstein and Suslow (2006, 2011)]. This success in colluding is partly because cartels develop structures to add flexibility and resolve challenges, thereby enhancing their prospects of survival. These include using compensation schemes, monitoring mechanisms to detect and deter cheating, and raising entry barriers [Levenstein and Suslow (2006)].

Some questions remain unanswered in terms of how cartels manage to collude and keep the incentive structure intact, especially in those circumstances where there is imperfect information/communication or where cartels have an imperfect mechanism for collusion (e.g. collusion on list prices or a part of the price mechanism). The answers to these questions may lie in the institutional structure of cartels and the compromises that they might make to survive.

Though the impact of multi-product/-market contact on collusion has been examined in the theoretical and empirical literature, the focus has been on single product collusion. Kovacic, Marshall and Meurer (2018) argue for the need to advance the theoretical literature beyond this, to examine multi-product/-market collusion as comprising cartels that are managing a portfolio of cartels rather than separate, individual cartels. The theoretical and empirical literature rightly acknowledges that multi-product/-market collusion can make collusion more likely and enhance its durability and stability. There is, however, limited work on other aspects of multi-product/-market collusion – for example, the pricing dynamics of multi-product cartels, and the nature and depth of harm resulting from such cartels. This is important for understanding collusion and its impact, and to increase the prospects of ensuring deterrence.

The widespread nature of collusion in South Africa justifies a proactive and aggressive approach to cartel enforcement. This approach has contributed to the detection of an expansive number of cartels and to penalties exceeding R7 billion. On the face of it, this suggests that cartel enforcement has been effective. However, measuring the effectiveness of cartel enforcement requires an appreciation of the full scope of collusion, and interventions that are geared towards effective deterrence. Effective cartel enforcement requires deterrence of the most durable cartels, some of which go undetected. One question arising from the cartel enforcement record relates to its efficacy. Answering this question requires a comprehensive,

critical examination of the cartel enforcement record and the resulting outcomes, from the perspective of deterrence.

This dissertation explores some of the unresolved issues through three key research questions, briefly discussed below. The goals of this dissertation were: (a) to examine, at a macro level, the characteristics of South African cartels and the institutional features of key selected South African cartels. In doing so, the author examined the extent of their consistency with, and the novelty of their features compared to the mainstream theoretical and empirical literature. This has important policy implications for cartel enforcement; (b) to examine the pricing behaviour of multi-product firms involved in collusion, with a view to estimating overcharges on different products and over time; and (c) to critically examine the cartel enforcement record, to evaluate its efficiency and effectiveness in achieving deterrence, and to make a case for an optimal cartel enforcement policy.

All three questions are important from an empirical and policy perspective. From a policy perspective, the study contributes to the design and implementation of cartel detection and enforcement tools. The study also contributes to a more effective policy approach to dealing with multi-product/-market collusion, which was prevalent among the detected cartels. In addition, it contributes to the empirical methods used to estimate cartel overcharges under conditions where there are temporal variations in overcharges – that is, where the cartel involves multi-product collusion, and where the price effects may dissipate in different ways at the end of the cartel. From a cartel enforcement perspective, the study contributes to the body of literature available on evaluating the efficacy of cartel enforcement, by collectively examining the drivers of cartel enforcement, the duration of cartel enforcement activities and their impact on penalties and deterrence.

While the dissertation did not directly aim to develop and propose new industrial organisation theory on collusion, some of its findings have theoretical implications in that they point to areas for further theoretical advancements. For example, given that cartels are found to use complementary forms of communication, including those that are not centralised to the cartel structure, the study proposes the need for advances in theoretical literature that incorporate these complementarities. These complementary forms of communication help cartels reduce uncertainties about the behaviour of cartel members and enhance the effectiveness of monitoring in cartels thereby contributing to the success of collusion. As such, incorporating these complementarities into theory will result in theory that is geared to explain collusion

under circumstances that existing theory does not fully explain. This is consistent with how theory develops, starting with observed firm behaviour which then gets incorporated into existing theory as an extension or incorporated into the development of new theory.

1.2.1 Research question 1: what are the characteristics of detected South African cartels, and the institutional features of selected key cartels?

The study reviewed the cartel record using case decisions from the competition authorities and other public information relating to the case studies of key selected cartels. The purpose of this was to examine the characteristics of South African cartels and the institutional features of selected cartels that have been detected in South Africa. In doing so, the study focused on how selected cartels managed themselves internally and how they managed to limit the external threats they faced. In answering this question, the author also examined the extent to which the cartel management structures were consistent with theoretical expectations and the findings from empirical studies. In addition, the study provides insights on some of the unresolved theoretical and empirical questions relating to collusion. The characteristics of cartels include the duration of firms' participation in collusion during the years covered by the current competition regime, the mix and forms of collusive practices employed by cartels, the extent of firms' involvement in collusion, and the extent of cartel overcharges.

The author found that the institutional structures of selected South African cartels were broadly consistent with those discussed in the theoretical literature and with the findings of empirical studies on cartels. Some of the important aspects of the findings relate to communication, the vertical structure of some selected key South African cartels and the role that protectionist trade tools might play when used as a strategic, costless tool to limit external threats. Communication plays a central role in collusion, but it can be imperfect, being neither formalised nor centralised by the cartel. Cartels are likely to use a combination of various forms of communication, and monitoring can comprise decentralised, firm-level forms, which nonetheless enhance collusion.

The theoretical literature does not always incorporate the combined use of various forms of communication and monitoring, which may also not always be readily reported in public information. For example, settled cases may not always expose other forms of communication and monitoring used by cartels, e.g. the use of matching clauses and the use of firm-level sales infrastructure and customer relationships. This could explain the difficulties faced in the theoretical literature when explaining the success of collusion in certain instances, such as

6

collusion on list prices when discounts are not agreed upon. The sales infrastructure of firms can provide a reliable way of detecting cheating, helping a cartel to deal with the uncertainty created by imperfect information.

Key South African cartels had a vertical structure to them, which likely helped them address both internal and external threats to collusion, in that some of the major cartel members were also suppliers of key inputs to the firms that they colluded with. In a country like South Africa, where the next best alternative source of inputs might be imports, trade tariffs also seem to be used as a costless tool for mitigating the threats that cartels face from imports.

1.2.2 Research question 2: what is the pricing behaviour of multi-product firms involved in collusion, and what is the extent of harm that they impose on consumers?

To answer this question, the study used panel-data analyses to examine the pricing behaviour of a multi-product cartel, with a view to estimating overcharges made by such a cartel on different products and over time. This question is important given the significant number of cartels involving multi-product/-market firms and the gaps in the empirical literature on these two aspects. The dissertation also contributes to the body of empirical literature on overcharge estimation when the dissolution of a cartel takes different forms.

The author found that cartel overcharges vary by product and over time, even if the collusive mechanism might be the same for different products. This finding suggests that harm cannot be generalised, and to the extent that penalties are in part influenced by the degree of harm imposed by a cartel during its life, there is a need to consider the pricing of different products over time by multi-product cartels. The variability of overcharges also means that in follow-on damages claims, the right approach would be to consider the harm pertaining to individual products bought by customers over time. The findings of this study could enhance cartel enforcement policy on penalties and deterrence. Taken together with the first research question, the study points to the need for closer scrutiny of multi-product/-market firms involved in collusion on one product or in one market, as the same firms may be involved in serial collusion.

1.2.3 Research question 3: what have been the key drivers of cartel enforcement in South Africa, and how effective has cartel enforcement been when examined through the lens of efficiency and deterrence?

Given the findings on the institutional features of collusion in South Africa and the pricing behaviour of multi-product firms involved in collusion, as well as the empirical evidence of overcharges from other studies,[2] it is important to provide a critical appraisal of the cartel enforcement record in order to understand cartel detection and prosecution through the lens of deterrence. The availability of resources, leniency and penalties have been the major drivers of enforcement activity, but they have not led to greater efficiency in cartel enforcement, as shown by the increasing duration of investigations and prosecutions over time. Panel regression results support the qualitative analyses of the duration of cases.

Using the present-value technique, the author examined the impact of the time taken to conclude investigations and prosecutions of firms on the present value of the penalties imposed. The study found that prolonged investigation and prosecution of cases had the significant effect of reducing the real value of the penalties paid. This undermines the efficacy of penalties as a deterrence tool. The key policy implication is that the Commission should account for delays in finalising cases, even if settled, when determining penalties, in order to preserve the deterrence effect of penalties. The author argues that an optimal cartel enforcement policy will penalise firms that delay finalising cases, with progressively higher penalties, in order to retain and enhance incentives to settle cases earlier, rather than later. In addition, interest can be added to the amount of the penalty for the period between the year on which the penalty is calculated and the year in which the penalty is paid.

1.3 Structure of dissertation

Chapter 2 presents an examination of the characteristics of detected South African cartels, focusing on the institutional features of selected key cartels. The pricing behaviour of a multi-product cartel is examined in Chapter 3, while Chapter 4 provides a critical examination of the cartel enforcement record over the 20 years from the start of the current competition law regime in 1999. Key conclusions and policy recommendations are discussed in Chapter 5.

---

[2] See for example, Mncube (2014); Khumalo, Mashiane and Roberts (2014)]; Mondliwa and Das Nair (2019); Boshoff (2015) and Boshoff and Van Jaarsveld (2019).

Chapter 2

Characteristics of South African cartels, focusing on institutional features of key cartels

2.1   Introduction

Collusion has been pervasive, occurring across several sectors of the South African economy. Some firms (either at the product level or at the group/holding company level) are involved in collusion across markets or sectors. As elaborated in Chapter 4, leniency played an important role in the detection and prosecution of cartels in South Africa. Leniency appears to have peaked in its ability to help detect cartels in South Africa, with applications having declined to pre-2010 levels (see Figure 2.1). This decline is consistent with the experience in other jurisdictions. In the European Union, leniency is also trending downwards, with 46 applications in 2014, 32 applications in 2015, 24 applications in 2016 and 18 applications in 2017.[3] The United States shows similar trends which have been attributed to the potential exposure to follow-on civil lawsuits.[4]

Figure 2.1: Number of leniency applications in South Africa from FY2000 to FY2018
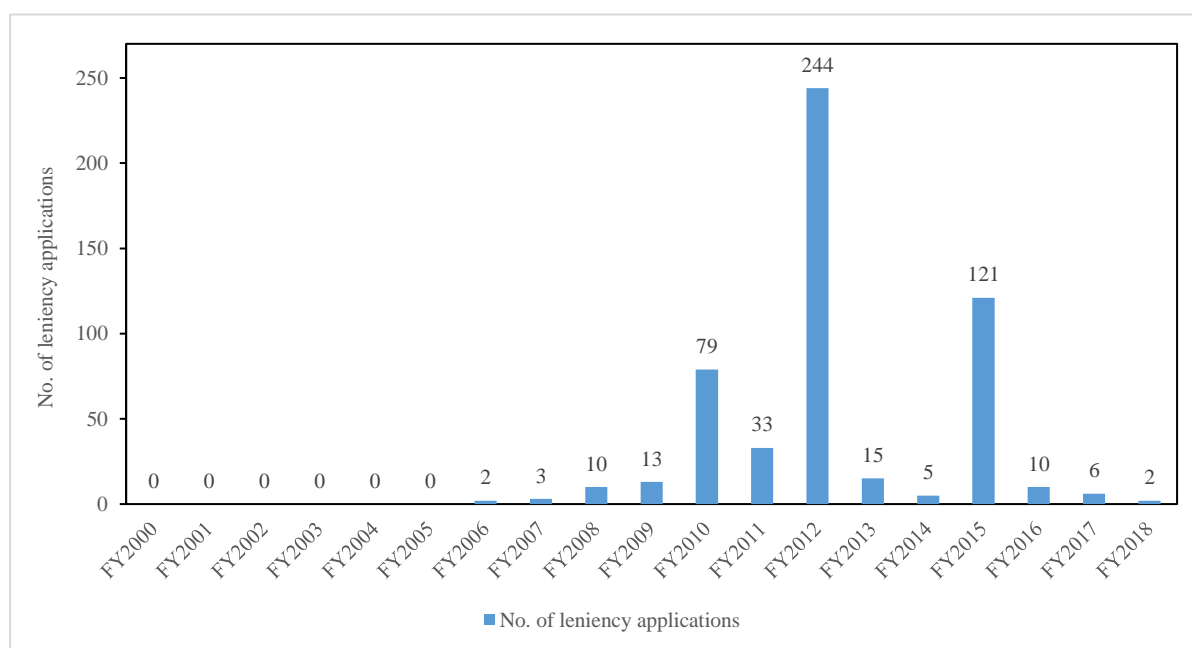
Source: Commission's Annual Reports

---

[3] https://www.cov.com/-/media/files/corporate/publications/2018/02/the_decline_and_fall_of_the_leniency_programme_in_europe.pdf and http://digital.shearman.com/i/1146712-shearman-sterling-antitrust-annual-report-2019/49?_ga=2.268162121.693788507.1591224999-1315488347.1591224999.

[4] http://digital.shearman.com/i/1146712-shearman-sterling-antitrust-annual-report-2019/49?_ga=2.268162121.693788507.1591224999-1315488347.1591224999.

As leniency applications have declined, and with concerns that leniency detects dying cartels,[5] the success of cartel enforcement in the future will depend on the ability of the competition authorities to proactively screen and detect cartels. Proactive detection will require the application of screening tools that rely on economic data and structural analyses of the markets.[6] Several factors are conducive to collusion, including high concentration levels; product homogeneity; the presence of high entry barriers; stable demand; firm symmetries; lack of excess capacity; high levels of market transparency, supported by information exchange; and the size and concentration of buyers. In a concentrated economy like that of South Africa, screening cartels solely based on these features might not be of much effect.

This chapter examined the key features of collusion in South Africa, particularly the institutional features of detected cartels.[7] The aim is to examine the extent to which the institutional features of South African cartels are consistent with the theoretical and empirical literature. This study is important for understanding how detected cartels in South Africa managed to keep the incentives to collude sufficiently compatible among their members for several years, and some of the strategic actions used to restrict external threats to the cartels.

The features of cartels detected in South African were drawn from a review of 265 prosecution decisions taken against firms between April 1999 and March 2018. Next, the study focused on 20 cartels to examine their institutional features. These cartels were selected based on their significance in terms of the potential impact they had on the economy or their institutional features, which provided insights into the management of collusion. Most of the cartelised products were inputs into other economic activities, while others involved important basic consumer products (e.g. bread and maize meal). These cartels were, on average, durable, operating for several years during their lifespan within the study period. The cartels included the:

(i)   flat steel cartel,

(ii)   long steel cartel,

(iii)   scrap metal cartel,

(iv)   rebar cartel,

(v)   wire cartel,

---

[5] Harrington Jr. and Chang (2015) and Harrington Jr. (2016).

[6] See Harrington (2015) for a discussion on both behavioural and structural screening.

[7] This is not to suggest that factors that are conducive to collusion were not relevant to the success of the cartels. They may well have been some of the key ingredients.

(vi)   cement cartel,

(vii) bread cartel,

(viii) milling cartel,

(ix)   construction cartels,

(x)   soda ash cartel,

(xi)   print media cartel,

(xii) ocean transport cartel,

(xiii) industrial waste-removal cartel,

(xiv) poultry breeding cartel,

(xv) chemical-to-fertiliser cartel,

(xvi) fishing cartel,

(xvii) wooden products cartel,

(xviii) glass cartel,

(xix) tyre manufacturers cartel, and

(xx) electrical cables cartel.

Collusion in South Africa has mostly been studied at a micro level, either at a sector level – e.g. in the construction sector, by Ratshisusu (2014) – or at an individual cartel level, mostly as parts of studies estimating cartel overcharges. The World Bank (2016) gives a high-level view of collusion in South Africa, focusing on the nature of cartels (e.g. duration, serial collusion, number of firms involved and linkages across cartels) and the features of the cartelised markets. However, the World Bank report did not examine, in detail, how cartels in South Africa organised and managed themselves. Examining the institutional features of cartels helps us understand how selected South African cartels organised and managed themselves, considering the multi-year duration of most cartels. It also helps determine their consistency with the theoretical and empirical literature, and the extent of uniqueness of collusion in South Africa.

Understanding the characteristics of cartels and the institutional features of key cartels helps competition authorities in two broad ways. First, it highlights areas of focus when screening for cartel conduct. An example of this would be to complement behavioural and structural screening by searching for the presence of joint ventures and sub-contracting arrangements between competitors where they are likely to be used to distribute collusive rents and keep incentives compatible.

Second, an understanding of the characteristics and features of cartels assists the competition authorities by highlighting the types of evidence they require to develop a coherent case for collusive harm. For example, where cartels use compensation schemes, joint ventures, sub-contracting and toll-manufacturing arrangements, investigations will require gathering evidence on payments between competitors, and understanding the nature of the work for which compensation is provided or the nature and rationale for the sub-contracting work undertaken. For toll-manufacturing arrangements, key questions might include, for example, whether there are legitimate business case reasons (e.g. capacity constraints) and why firms favour toll-manufacturing over expanding their own capacities (e.g. cost justifications). When investigating collusion in markets where there are allegations of abuse of dominance, competition authorities should not treat such allegations as being isolated allegations against a single dominant firm. Instead, possible abuse of dominance should be incorporated into a coherent theory of collusion, in which the alleged abuse may in fact be an entry or expansion deterrence strategy or a tool for internal management of the cartel.

The chapter is organised as follows: section 2.2 provides a discussion of the theoretical and empirical literature on collusion, focusing on the institutional structures of the cartels; section 2.3 highlights the key features of collusion in South Africa in light of theoretical and empirical literature; section 2.4 provides a discussion of the key institutional features of cartels, using case studies of some of the key cartels in South Africa; and key conclusions and policy implications are discussed in section 2.5.

2.2   Review of literature on theory and empirical evidence of collusion

2.2.1 Brief review of theoretical and empirical literature on the institutional features of cartels

Collusion requires aligning incentives and managing the self-interests of competing cartel members (i.e., the incentive to cheat on the collusive scheme), and managing the threat of external destabilising factors. These are not always easy to achieve. This section focuses on the theoretical and empirical literature that examines the institutional features of cartels, focusing on those that influence their internal and external stability.

2.2.1.1 Internal institutional features of cartels

In his influential paper, Stigler (1964) concluded that collusion is inherently unstable, as the private incentive to cheat on the collusive arrangement will prove powerful enough to outweigh the benefits of co-ordination to individual firms. This led Stigler to conclude that cheating will be the primary cause of the death of cartels. In any given cartel, the incentive for firms to unilaterally defect exists for several reasons. A cartel member facing financial challenges may have a different intertemporal discount rate compared with other firms, and this may mean that the firm will lack the patience to honour the collusive agreement, leading to cheating [Porter (2005) and Levenstein and Suslow (2011)]. Cartel members facing different costs or capacities may also resort to cheating [Cave and Salant (1995); Athey and Bagwell (2001); and Athey, Bagwell and Sanchirico (2004)]. A maverick firm may also have incentives to cheat [Baker (2002)]. Dissatisfaction over the sharing of cartel rents could result in cartel members having different valuations of the collusive rents, leading to cheating [Hendricks, Porter and Tan (2008)].

This literature highlights the internal challenges that cartels face as they try to raise prices above competitive levels. However, it does not fully consider the various factors and ways in which cartels can, with some success, mitigate the challenges. Equally, the early theoretical literature often does not formally address entry as a threat to collusion, meaning that cartel instability was regarded to be a function of internal dynamics. Recent advances in the theoretical and empirical literature on collusion are useful in understanding how cartels organise themselves and manage the threats to their existence.

Porter (2005) developed a useful taxonomy of factors and institutional structures of cartels, which aid collusion. These factors and structures help align disparate interests, address cheating, and enable cartels to adapt to changing market conditions. Where cartels face disparate interests, they must find a way of accommodating members' interests. This requirement affects the collusive mechanism used by the cartel and how the cartel deals with instances of unilateral defection and changes in market conditions. For instance, in cartels involving large and small firms, the larger firms may have to make compromises with the smaller firms to co-opt them into the collusive scheme, rather than alienate them. This affects how the economic rents from collusion are shared between the large and small firms [Porter (2005)]. Some cartels use joint ventures and side-payments as mechanisms to address the varied interests of members [Porter (2005) and Levenstein and Suslow (2006, 2011)].

13

Communication is central to cartels resolving these issues, and as such, the next section reviews the literature on communication within cartels.

2.2.1.1.1 Communication and monitoring

Communication and information are important to cartels in two ways. First, they help cartels reach terms of co-ordination, including the collusive mechanisms and sharing of rents. Second, they help monitor the actions of cartel members, making implementation more effective. Albæk, Møllgaard and Overgaard (1997) give a useful example of how the availability of information can help oligopolies function better and more effectively in raising prices above competitive levels. The authors found that the prices of two grades of ready-mixed concrete increased by between 15 per cent and 20 per cent within a year of government publishing firm-specific transaction prices for those products. This shows that beyond facilitating agreement on terms of co-ordination, communication is an active part of the collusive equilibrium.

Communication between cartel members can be direct (e.g. through meetings, telephonic conversations, and emails) or indirect (e.g. using signals). The citric acid cartel in the US, for example, used structured cartel meetings to agree on prices, share information and resolve cartel problems. In the case of indirect communication, one cartel member could use advance price announcements to send a signal to other cartel members [e.g. Grether and Plott (1984) and Holt and Scheffman (1987)]. These forms of communication are imperfect, and to be effective, they must communicate the minimum level of information required to enable cartel members to co-ordinate.

Cartels try to eliminate problems caused by information imperfections by investing in monitoring mechanisms that can track the actions of individual members. These monitoring mechanisms evolve over time as cartels experience new challenges. Their degree and nature vary across cartels depending on each cartel's circumstances. Effective monitoring helps cartels to detect cheating and to separate cheating from demand fluctuations, enabling them to avoid unnecessary, costly price wars [see Levenstein and Suslow (2006)]. Cartels can share information (e.g. sales volumes) among themselves, as was the case in the Lysine cartel, or they can use trade associations or other independent third parties [see for example, Levenstein and Suslow (2011)]. In some cartels, the role of organisations such as trade associations extends beyond information sharing to a more active role in the management of the cartel. For example,

the Sugar Institute acted as an arbiter when there were disputes about cheating [Genesove and Mullin (2001)].

Monitoring cartel members through information exchange can be complemented by information from customers, e.g. through price matching clauses, which can help monitor defection by fellow cartel members [Porter (2005)]. The same applies to information that sales teams can glean from day-to-day interactions with customers. These forms of monitoring are driven by individual firms keen to protect their own share of the collusive rents, and they take advantage of attempts by unsuspecting customers to get better prices from suppliers. These types of monitoring mechanisms can be more effective in detecting and responding to secret price cutting, and help cartels avoid engaging in indiscriminate, costly price wars. Responses can be targeted, simply removing the benefit of cheating, thereby reducing the incentives of cartel members to cheat.

Information from the interactions between cartel members and their customers helps with monitoring the price side of firm behaviour, complementing monitoring done through sales reports exchanged between cartel members. The complementary use of the two helps cartels know when a decline in sales for some cartel members is the result of cheating or the result of a slump in demand. This helps cartels function better.

The economic literature identifies different forms of communication and different sources of information for cartels. However, it does not fully examine the feasibility of collusion when cartels use a combination of communication mechanisms to develop more complete monitoring mechanisms. This failure to recognise and account for complementary forms of communication and monitoring means that economic theory, in some instances, is unable to fully explain collusion when certain aspects of firm behaviour are not readily monitored in a centralised way. In this chapter, the author attempts to develop further understanding of collusion, by examining communication within South African cartels, with a focus on selected key cartels. In so doing, the author assesses the extent to which cartel experiences align with or deviate from the literature.

The next section provides a review of the economic literature on how cartels address cheating, which has been identified as a prime threat to collusion.

2.2.1.1.2 Cheating

Cheating happens in most cartels and, if left unchecked, can result in the collapse of a cartel. Cartels look for ways to deal with cheating by their members, and there are several strategic actions that firms can adopt to help deal with this problem. Cartel members could, for example, use a price war to make cheating unprofitable. Green and Porter (1984) conclude that the threat of a costly price war deters cheating. By extending the work of Green and Porter (1984), Abreu, Pearce and Stacchetti (1986) identified a set of strategies that maximise the expected profits from collusion with imperfect monitoring. They conclude that on-and-off price wars do not reflect unstable cartels, as posited by Stigler, but are mechanisms through which cartels manage themselves.

Complex pricing structures make it difficult to detect cheating. Some cartels solve this problem by simplifying pricing structures. Examples of simplified pricing structures include standardised pricing in the case of the Sugar Institute, which Genesove and Mullin (2001) studied, and base-point pricing structures. Cartels use the simplest available mechanisms given the circumstances they face – for example, the distribution of market shares, products or geographical areas, or customers among members – and leave the pricing decisions to individual firms.

Some cartels use compensation schemes that involve payments or transfers among cartel members [Porter (2005) and Levenstein and Suslow (2006, 2011)]. These disincentivise cheating, by shifting the illicit gains of defectors to the victims of defection. This reduces the incentive to cheat. In turn, this keeps the agreed framework of rent sharing in balance.

The fact that firms in a cartel typically interact repeatedly creates greater scope for co-operation because cheating in the current period can induce aggressive responses from other cartel members in future periods. The seminal work of Friedman (1971) complemented by research such as by Benoit and Krishna (1985) shows that collusion can be supported by repeated interaction over time and across markets, which makes punishment more effective in deterring cheating, especially where firms value future profits more than current profits. Repeated interaction also produces learning effects in cartels and, where possible, scope for refinements to the cartel's organisation, e.g. by enhancing information gathering (monitoring) capabilities. This increases the prospects of successful collusion in terms of duration and ability to set prices above competitive levels.

16

Harrington (1987) studied the possibility of collusion in finitely repeated games under the assumption of full rationality and found that there were instances where collusion could be achieved. The author also found that such collusion could be extended to a multi-product/-market setting where future behaviour in one market becomes dependent on behaviour in another market, increasing the likelihood of reaching a cartel solution. Multi-product/-market collusion could enhance the credibility of punishment being extended to other products by raising the cost of defection.

Bernheim and Whinston (1990) similarly found that multi-market contact could enhance collusion and benefit the firms involved, but that the impact of multi-market contact was not necessarily socially undesirable. The authors argue that the impact on prices or profits in any of the affected markets depends on the nature of the markets and the features of the firms involved, and on potential entrants. When there are cost asymmetries between firms, discount rates will influence whether prices are lower or higher. If firms are similar, prices may be higher in some markets but lower in others. The authors argue that with multi-market contact, behaviour in different markets/products becomes interdependent.

Spagnolo (1999) identifies circumstances in which multi-market contact always facilitates collusion – for instance, when strategic interactions between firms are interdependent, meaning that a firm's valuation of profits in one market is influenced by profits achieved in other markets. In this situation, punishment is more effective, and collusion is more profitable when multi-market contact exists. Matsushima (2001) shows that multi-market contact facilitates collusion even under imperfect monitoring, especially if there is interdependence in the supply choices of firms in the different markets. At least one study shows that firms that interact in multiple markets are more likely to co-operate in pricing behaviour than those that do not [see Ciliberto and Williams (2014), who studied collusion in airlines]. The authors, however, argue that the impact of multi-market contact on prices is influenced by cross-price elasticities.

Kovacic, Marshall and Meurer (2018) examined the incidence of serial collusion by multi-product firms, noting that the evidence reflects business models that embrace collusion. The authors argue that such collusion reflects the management of a portfolio of cartels rather than individual cartels. The authors conclude that serial collusion creates opportunities for cartels to solve the standard cartel problems more effectively, making collusion more profitable, stable, and long-lasting. Cartels involved in serial collusion can better share collusive rents across multiple products, punish defections more effectively across product lines, and have a wider

set of possible collusive equilibria to enhance the success and profitability of the overall portfolio of cartels. Kovacic, Marshall and Meurer (2018) argue that theories based on single-product collusion do not capture the full dynamics of collusion in a serial setting. This in turn leads to limited appreciation of the nature and effectiveness of collusion, limitations in the detection of collusion, and enforcement efforts that are not fully effective in deterring collusion. Often competition authorities consider cartels at a market/product level, meaning that cartel stability and harm are dealt with from the perspective of individual markets/products, as opposed to the portfolios of cartels that firms may be involved in.

The review above shows that cartels have various mechanisms for deterring cheating. These include price wars, which can be in the form of a symmetrical shift to non-collusive behaviour by all members or asymmetrical targeting of only the cheating member. Simplified pricing structures remove the scope for cheating, and compensation schemes eliminate the benefit of cheating and restore agreed rent sharing structures. Repeated interaction across markets/products introduces flexibility and increases the cost of deviation. Similarly, repeated interaction over time also creates the scope for making punishment more effective, especially where firms value future profits more than current profits. Collusion in multiple markets means that firms have flexibility in sharing collusive rents and in resolving cartel problems.

Cartels can also take advantage of vertical integration or vertical relationships to enhance their prospects of success. The next section reviews the relevant literature in this regard.

2.2.1.1.3 Vertical integration and vertical arrangements

Vertical integration can, through enhanced information exchange, ease and increase the level of transparency in markets, helping firms reach and sustain collusion. Firms can achieve collusion using vertical agreements that enable information sharing among cartel members. This is the case with 'hub and spoke' cartels, where a supplier or retailer acts as the platform (hub) for sharing information between the cartel members (spokes) operating at another level of the value chain [Niels, Jenkins and Kavanagh (2011)].

Cartel members can also use vertical restraints (e.g. resale price maintenance, exclusive territories) to soften competition, by preventing aggressive competition at another level of the value chain [Niels, Jenkins and Kavanagh (2011)]. When suppliers are colluding, these kinds of arrangements can reduce both intra- and inter-brand competition at the retail level, helping sustain higher prices. Retailers can undermine a suppliers' cartel by indirectly promoting

competition between themselves through discounting the prices of the different suppliers. The suppliers' cartel could prevent this by employing restraints that prevent retailers from discounting the products of the various suppliers.

Arrangements that seem vertical in some respects (e.g. toll manufacturing between cartel members) can be employed by cartels to reduce competition and enhance rent sharing. For example, cartel members may opt for toll manufacturing arrangements over investing in expanding their own capacity. In so doing, the cartel members avoid increasing market capacity, which could incentivise competition. Some cartels may involve firms exiting one level of the value chain and simultaneously entering into a vertical supply agreement with a fellow cartel member for the supply of the upstream product that they had previously supplied themselves.

Although the literature identifies the potential use of vertical integration and vertical restraints to support collusion, it does not address collusion in situations where vertically integrated firms participate in collusion at successive levels of the value chain. This study examined the vertical features of collusion in key South African cartels, particularly the extent to which vertical integration, vertical agreements, and collusion took place at successive levels of the value chain. The latter is not fully explored in the existing literature.

2.2.1.2 External threats of entry and expansion

In Levenstein and Suslow (2006), entry is found to be the most common cause of cartel death, while it is found to destabilise cartels in the Levenstein and Suslow (2011) study of international cartels. Cartels that succeed exercise cartel market power and raise prices above competitive levels, but such prices inevitably attract entry, which in turn erodes the supra-competitive profits. Entry will therefore undermine the collusive schemes. Cartels are more likely to survive and be effective in raising prices above competitive levels for a sustained period in markets characterised by significant entry barriers than in markets where barriers are low [see Dick (1996b); Symeonidis (2003); Levenstein (1995); and Clay and Werner (2002)].

Some cartels create barriers to prevent or marginalise entry by new competitors or the expansion of firms that are outside the cartel. These barriers include those of (i) exclusionary behaviour by the cartel or some of its members – e.g. in the form of predation [see Burns (1986); Weiman and Levin (1994); Scott Morton (1997); Podolny and Scott Morton (1999); and Levenstein and Suslow (2011)] – or using vertical integration or vertical restraints [Bishop

19

and Walker (2010)]; (ii) information asymmetries between incumbents and new entrants; and (iii) regulatory impediments imposed by the state on its own initiative, at the request or with the support of firms involved in cartels, e.g. the imposition of import tariffs and other restrictions on entry. The theoretical literature shows that strategic entry deterrence is possible where there is some form of excess capacity, which makes responses to entry or threats of entry more credible [e.g. Dixit (1979)].

Genesove and Mullin (2001), Porter (2005) and Levenstein and Suslow (2006, 2011) argue that resilient cartels embrace learning effects by adapting the underlying collusive agreement or implementing mechanisms that enhance their ability to cope with evolving circumstances, e.g. new ways of sharing information and making cartel decisions. Harrington and Chang (2009) found that cartels that cope with shocks are likely to last longer than those that do not. The response to entry and threats of entry is also an important part of a successful cartel operation.

2.2.1.3   Summary of key learning from literature

Cartels need to communicate to succeed. The form of communication varies, and different cartels use different forms of communication that, in some cases, involve extensive information sharing. The literature does not fully examine the impact of the combination of different forms of communication and monitoring by cartels, which could explain the success of collusion when some monitoring mechanisms are imperfect.

Price wars help deter cheating in some cartels, but other mechanisms can sustain collusion (e.g. compensation schemes, multi-market contact, and vertical integration and restraints). All these are recognised in the economic literature. However, there is limited consideration of collusion where vertically integrated firms participate in collusion at successive levels of the value chain. As section 2.4 shows, this is a feature of some of the key cartels in South Africa. Section 2.4 confirms the use of compensation schemes and the correlation between multi-market contact, vertical integration and restraints, and collusion.

The institutional features of collusion can help during the screening and investigation of suspected cartels, especially in an environment where the primary detection tools such as leniency have deteriorated. Some studies [e.g. Levenstein and Suslow (2006,2011) and Porter (2005)] have considered the factors that help cartels succeed, including institutional features. This chapter conducted one such study, which has not hitherto been done for South Africa.

Section 2.3 provides a discussion of the broader dataset used in the study and then proceeds to discuss the characteristics of collusion in South Africa.

2.3 Characteristics of collusion in South Africa

This section provides a discussion on the characteristics of those cartels that were detected and penalised in South African between April 1999 and March 2018. The focus was on the forms of collusion used by the cartels, the extent of firm involvement in collusion across sectors, the duration of firm involvement in collusion, and evidence of the pricing power of some of the detected cartels. The data and its sources used in this chapter are described, before discussing the characteristics of the detected cartels.

2.3.1 Description of data and its sources

The data used in this study came from the cartel prosecution records reflected in the decisions of the Competition Tribunal (Tribunal) and the Competition Appeal Court (CAC) covering the period from April 1999 to March 2018, comprising 265 prosecutions of firms across 8 sectors. This data was generated from the investigations and prosecutions of firms by the Commission and is reflected in the case decision records that describe the cartelised products, the nature and features of the conduct being prosecuted, the firms involved, whether the investigation involved leniency, and how the prosecutions were resolved. The data on cartel features typically includes the type of conduct (e.g. price fixing, market allocation, cover pricing, collusive tendering), when the conducted occurred, other elements of the collusive arrangement (e.g. joint ventures, payment schemes and sub-contracting) and in some instances the involvement of an industry body (e.g. industry association or third-party provider of information).

Most cases (over 90 per cent) prosecuted between April 1999 and March 2018 were concluded by way of settlements, in which the Commission and the firm under investigation or prosecution negotiated the terms of settlement, principally the penalty paid by the firm and the admissions made on the collusive conduct. In many instances, the firms pledged to help the Commission in its investigation and prosecution of other firms involved in the cartel or in relation to collusion involving other products or markets that the firm was involved in. Unsurprisingly, litigated cases yield much richer information than settled cases.

The information was obtained from the Tribunal's and the Commission's websites, annual reports, press releases and other publications. Some cartels had been in operation since prior to 1999, but the study excluded this prior period because it fell outside of the applicability of the current Competition Act, which introduced more vigour in cartel enforcement in South Africa. The period prior to 1999 also coincided with State regulation of markets, in which some of the key cartels were either legalised (e.g. cement) or part of marketing boards (e.g. the grain-milling cartel). This history was, however, useful for understanding an aspect of the context of the illegal cartels under the current competition law regime.

Given the way in which prosecutions were mostly resolved through settlements, the data presented some limitations. First, the data was likely to be incomplete because cases that were concluded via settlements and leniency did not always result in full exposition of a cartel's institutional structure, its evolution, and how the cartel managed the challenges it encountered during its life. For example, in the publicly available cartel data, there was limited investigation and discussion of how the cartels dealt with the threat of, or actual entry or how they might have dealt with the presence of a competitive fringe. The data was complemented using other data relating to the activities of the firms in other spheres such as policy or regulation-making processes, the incidence of complaints against cartel members for exclusionary conduct, and the like. Similarly, there was not always full discussion of the institutional structure of cartels, to understand how those structures helped collusion, or what challenges the structures in place resolved.

The second limitation relates to the sample, which was biased towards detected and prosecuted cartels, and the observed features of these cartels may have differed from the features of durable undetected cartels. Nonetheless, the dataset provided some rich information on how cartels in South Africa organised themselves and sustained their arrangements. These two broad limitations of the data are not unique to this study and are unlikely to invalidate the insights drawn, because the available baseline data describes the key features of the collusive arrangements in the prosecuted cartels.

2.3.2 Forms of collusion

In South Africa, collusion is prohibited along three main dimensions: price fixing; market allocation and collusive tendering. Within these three broad dimensions, various types of conduct are captured. For example, the prohibition of price fixing covers direct and indirect

price fixing and the fixing of other trading conditions. This means that even fixing discounts is covered and as such prohibited. Market allocation covers the allocation of customers or suppliers, geographic territories, specific types of products and services, and market shares. Collusive tendering is a broad category that covers various forms of collusion in relation to tenders. For example, cover pricing would be covered under collusive tendering.

From an economics perspective, the form of collusion is not necessarily the most important consideration leading up to its prohibition, but the effect that it has on pricing. Firms collude to raise prices above competitive levels so that they earn supra-competitive profits. This can be achieved in different ways, depending on the circumstances faced by the cartel, including agreeing on the price; allocating customers to cartel members, and leaving the pricing to cartel members; or agreeing on restricting output to a certain level, which has the effect of raising the price; or even agreeing on the level of profit. The available options for collusion are not mutually exclusive, and cartels can use them in combination. For example, a cartel that faces a variety of customers who buy products using different mechanisms, such as normal day-to-day transactional purchases and tenders, is more likely to collude across the three main dimensions. For the group of customers who use tenders, the cartel might opt to use cover pricing or customer allocation, and for customers who use day-to-day purchase systems, the cartel might fix prices, discounts, or both.

The cartel enforcement record shows that of the firms that were prosecuted between April 1999 and March 2018, 31 per cent were involved in a cartel that used a combination of different forms of collusion. From his review of collusion in the construction sector, Ratshisusu (2014) noted that firms used customer (bids) allocation together with agreements on the profits to be earned from contracts. The same observations can be made of the fertiliser cartel, which used market allocation together with an agreement to keep target margins. In the wire industry, the cartel fixed prices, agreed on discount levels for different classes of customers and allocated some customers among cartel members. The milling cartel fixed prices and changes in prices for different products and classes of customers, and allocated customers among cartel members. Similarly, firms involved in the bread cartel fixed prices, increases to the prices of some types of bread and discounts or commissions given to agents, and allocated customers. The cement cartel employed price fixing coupled with an agreement not to offer special discounts on higher quality cement, market share allocation and geographic market allocation.

Firms that colluded in the scrap metal industry used price fixing (selling and purchase prices for scrap metal) and customer allocation as mechanisms for collusion.

69 per cent of the prosecutions related to cases where the cartels employed only one form of collusion. Of this group of prosecutions, 55 per cent were for price fixing alone, 36 per cent were for collusive tendering and 9 per cent for market allocation. There was no obvious bias towards specific economic activities in the use of each of these forms of collusion. This is likely because the form of collusion was driven by the nature of the customer purchases (e.g. the use of tenders) and within that, the easier co-ordinating practice, rather than the nature of the economic activity. Although price fixing was widely used, it was not frequently used in construction and transport (furniture removals for instance), where customers mostly used tenders to procure services.

2.3.3 Extent of firm involvement in collusion across sectors

Firms collude because it is more profitable to do so than to compete, as suggested in studies of cartel overcharges (see section 2.3.5, below, for some examples). Collusion has been a pervasive problem in South Africa, demonstrated by the extensive record of prosecutions of firms in the country, spanning various sectors, with some companies (either at a product level or at a group/holding company level) being involved in collusion across markets or sectors. Estimates from the data described in section 2.3.1 show that 42 per cent of firms prosecuted between April 1999 and March 2018 colluded more than once or belonged to a group in which another firm was prosecuted for collusion. Figure 2.2 shows the extent of firm involvement in collusion across sectors, based on the enforcement record of case decisions taken by either the Tribunal or the CAC in South Africa from the start of the modern-era South African Competition Act in 1999 until March 2018. Firms have been prosecuted for collusion in agriculture, communications, construction, health, manufacturing, services, trade, and transport, with the distribution shown in Figure 2.2, below.

Figure 2.2: Extent of firm participation in collusion across sectors
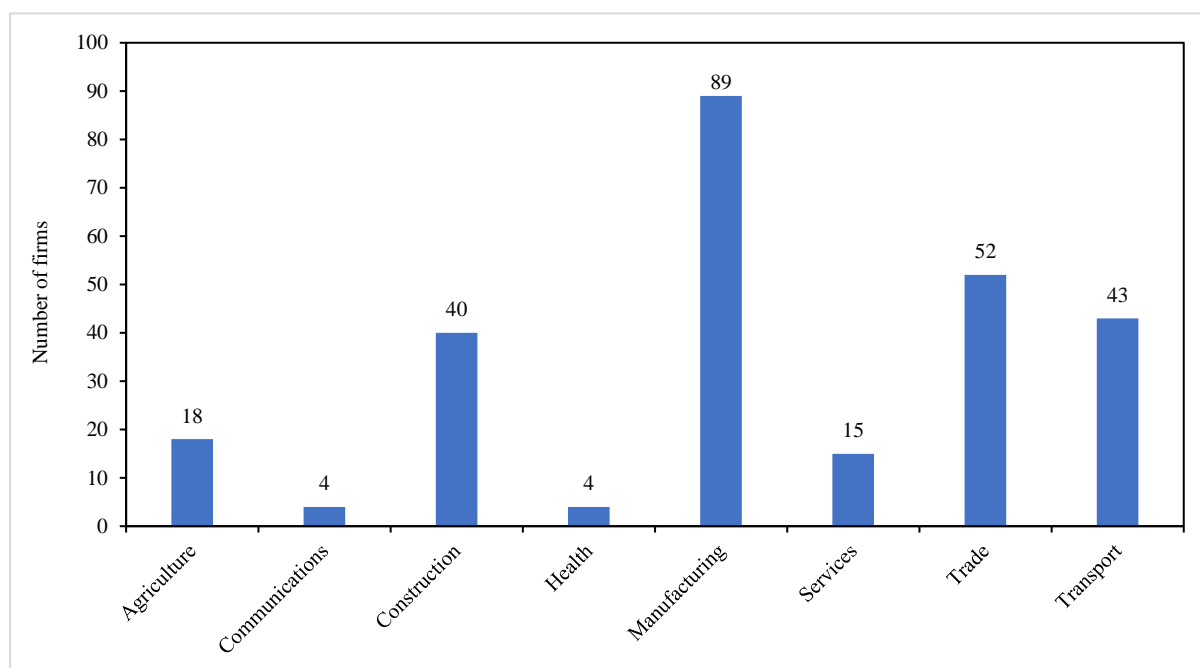
Source: Own compilation from various case decisions of the Tribunal and CAC

The number of firms per cartel varies widely, ranging from just two[8] to as high as 20 in cartels such as the bicycle retailing and rebar cartels. The maize milling cartel comprised 17 firms. The large number of firms in some cartels reflects the fact that to be able to raise and sustain prices above competitive levels, a cartel must control a significant amount of output. In markets with many firms and where firms can expand their output, this can only be achieved by incorporating as many firms as possible into the cartel; otherwise, attempts to raise and sustain prices above competitive levels for a sustained period would be undermined by those outside it.

In other instances, despite the large number of firms involved in a cartel, most of the output was controlled by a few firms. The inclusion of smaller firms in those cartels could reflect accommodation strategies by larger firms or industries characterised by few leader firms and small follower firms, with limited capacity to aggressively challenge the larger firms. Rather than being outside of cartel arrangements and adopting a challenger position, the smaller firms may find it better and more profitable to be part of the cartels. This is more likely to be the case in those markets characterised by some form of economies of scale and scope where fixed costs are significant, and where larger firms have a per-unit cost advantage and, as such, are better

---

[8] Note that due to the *per se* nature of the prohibition of cartel conduct, firms can be prosecuted even if they do not have the ability to raise overall market prices above competitive levels. The mere agreement on a price leads to prosecution.

placed to aggressively deal with a smaller firm that adopts a maverick stance. As such, by being part of a cartel, the smaller firms are shielded from the negative effects of aggressive competition which might push them to exit the market. Markets such as maize milling in South Africa had a history of co-operation, and it is possible that all firms just continued with the more comfortable tradition of cooperation after 1999, rather than moving to a new tradition that embraced a riskier competitive culture.

2.3.4 Duration of firm involvement in collusion since 1999

The duration of a cartel and of firms' involvement in collusion provides a useful indicator of the stability of collusion and of the success of collusion, especially where there is evidence that the cartels were able to raise and sustain prices significantly above competitive levels for an extended period. The World Bank (2016) gives some useful facts about some of the cartels that have been detected and prosecuted in South Africa. For instance, in a review of cartels prosecuted between 2005 and 2015, the World Bank notes that on average the cartels ran for eight years.

In the database of firm prosecutions between April 1999 and March 2018, this study estimated the average duration of involvement of firms in collusion. This was done in two ways: (i) across economic activities and (ii) for firms involved in serial collusion or that belonged to a group in which another firm was prosecuted for collusion relative to those that were involved in only one cartel and were not part of a group in which another firm was prosecuted for collusion. For the reasons stated above, duration of involvement was calculated from 1999, even though for some cartels, collusion started prior to this.

For firms involved in collusion that involved bidding for contracts, there can be debate on how duration should be determined, that is, whether in terms of the specific contracts or in terms of the serial nature of collusion across multiple simultaneous and sequential contracts. In this study, the author argues that the duration of involvement should not be determined based on the duration of a specific contract over which firms cover-priced or allocated customers. Using the duration of a single contract does not yield appreciation of the *quid pro quo* nature of the collusive arrangements in their totality and fails to consider the incentives of firms to provide a cover-price on one contract when they benefit from another contract. In such cases, it is highly probable that firms gain from the other contracts that they get around the same time or when their turn comes in a sequential bid. This is akin to the observations of Kovacic, Marshall and

26

Meurer (2018) on the benefits of involvement in a portfolio of cartels. A review of collusion in construction cartels shows that the *quid pro quo* nature of collusion applied to a series of contracts taking place in the same period or in different periods. In the author's view, duration should thus be from the beginning of the first instance of collusion in the series of cartelised contracts to the end of the cartel or the firm's participation in collusion.

In the study sample, a firm was involved in collusion for an average of 6 years, with the shortest duration being less than a year and the longest being 17 years.[9] Figure 2.3 shows the average duration of firms' involvement in collusion by economic activity[10], as well as the shortest and longest durations. Across various economic activities, collusion lasted for several years, which is consistent with observations made by the World Bank (2016).

Figure 2.3: Average duration of firm involvement in collusion from 1999 to 2018
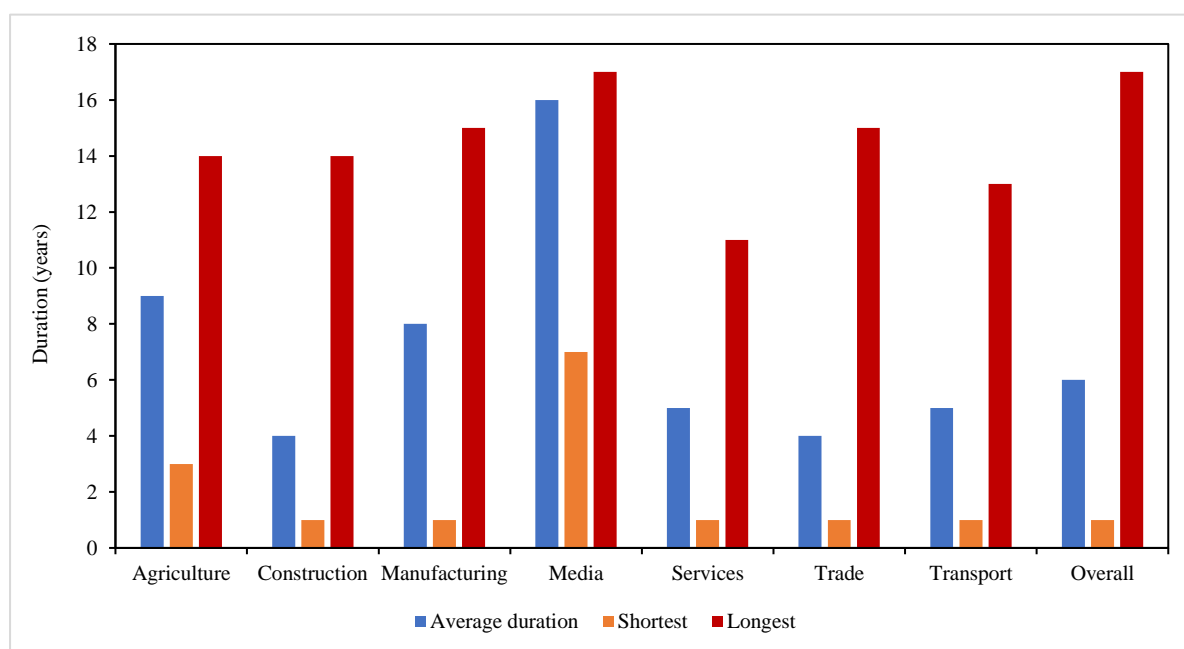
Source: Own compilation from various case decisions of the Tribunal and CAC

Figure 2.4 compares the average duration of involvement in cartel conduct by firms that engaged in serial collusion and those that did not.

---

[9] The average duration of firm involvement in a cartel can differ from the average duration of a cartel because some firms join cartels that are already in existence and may exit cartels that are continuing.

[10] Economic activities under each broader economic activity group were varied in nature. For example, manufacturing involved a wide variety of activities across different markets.

Figure 2.4: Average duration of involvement in collusion among serial colluding firms and once-off colluding firms
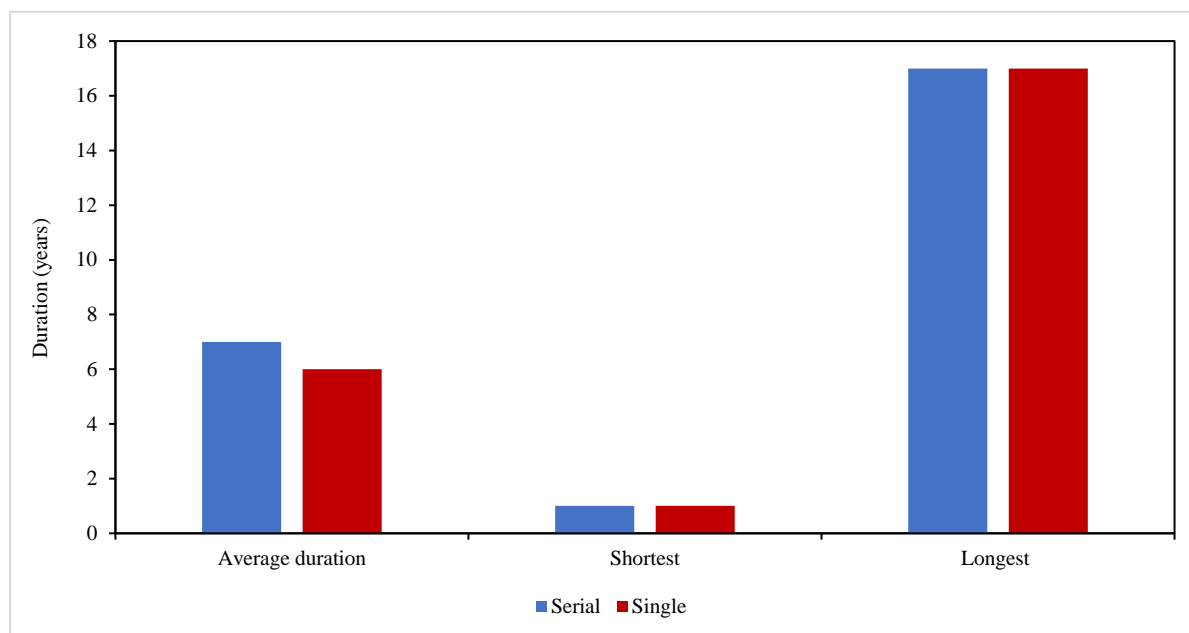
Source: Own compilation from various case decisions of the Tribunal and CAC

Concerning collusion arising from serial involvement or being part of a group where one of the sister firms was involved in collusion, there did not appear to have been major differences in the duration – other than by one year longer – of these firms' involvement in collusion. This result was due to some once-off colluding firms being involved in the same cartels as serial colluding firms and staying in these cartels for the same duration.

2.3.5 Evidence that cartels exercise market power afforded by collusion

A wide range of international and South African empirical studies on cartel overcharges show that many cartels were able to exercise collective market power facilitated by collusion, charging prices that are significantly above competitive levels.[11] The degree to which individual cartels succeed in raising prices above the competitive levels varies by cartel and product. Connor and Lande (2006) found that average cartel overcharges varied significantly across empirical studies, with average overcharges ranging between 31 per cent and 49 per cent and the median overcharges ranging between 22 per cent and 25 per cent. Werden (2003) estimated average overcharges of 21 per cent, while the OECD (2002) estimated average overcharges of 15.75 per cent. Using 12 cartel cases, Posner (2001) found average overcharges

---

[11] It is important to note that the results are sensitive to estimation procedures employed and the quality of the data used in the estimation procedure.

28

of 49 per cent. Oxera (2009) found median overcharges of 18 per cent and mean overcharges of 20 per cent. In a study of overcharges in the German cement cartel, Hüschelrath, Müller and Veith (2013) found overcharges ranging from 20.7 per cent to 26.5 per cent, depending on the estimation approach.

There have been several studies on cartel overcharges in South Africa. A few illustrative examples show how much cartels in South Africa were able to exercise collective market power. Mncube (2014) estimated overcharges ranging between seven per cent and 42 per cent for the wheat flour cartel. With the concrete pipes cartel, depending on the geographic region, overcharges ranged between 16.5 per cent and 57 per cent [see Khumalo, Mashiane and Roberts (2014)]. The reinforcing bar cartel achieved overcharges of 35.7 per cent and 43.2 per cent, depending on the counterfactual [Mondliwa and Das Nair (2019)]. In the bitumen market, Boshoff (2015) found overcharges ranging between 25 per cent, based on the forecasting and dummy variable approaches, and 80 per cent, based on the difference-in-difference approach. Boshoff and Van Jaarsveld (2019) found estimated cartel overcharges by the cement cartel, of between 19.2 per cent and 19.9 per cent.

Chapter 3 presents the findings of the study, which explored overcharges by a multi-product cartel that operated over multiple collusive phases. As shown in Chapter 3, average overcharges varied by product, with smaller overcharges, which are not statistically significant, for some products, and larger overcharges, which are statistically significant, for others. Overcharges were also found to have varied across collusive phases.

2.4   Institutional characteristics of key South African cartels

This section provides a discussion of the institutional features of the key cartels identified in section 2.1 above. The cartels were significant in terms of their potential impact on consumers and the economy, their institutional features, or their durability. The focus was on those features that related to the internal management of cartel arrangements and those that related to the management of external threats. The study made inferences and drew conclusions by contextualising the institutional features observed in the cartels in terms of economics theory and the empirical studies of collusion.

2.4.1 Internal management of collusion

Examining the internal institutional features of the selected cartels shed light on their durability, which lasted several years for many of them. The author was interested in two important elements of collusion, which are recognised in the theory and in practice. The first of these was the alignment of incentives, and the second, dealing with cheating in a way that was effective enough to keep incentives aligned for the duration of the cartel. Supra-competitive profits from collusion can create powerful motivation for firms to align the incentives to collude and to keep them sufficiently compatible to sustain the collusion [Levenstein and Suslow (2006)]. Estimates of cartel overcharges by South African cartels suggest that collusion creates the scope to raise prices significantly above competitive levels, yielding significant profits that can be sustained for several years.

Communication is important for firms in aligning incentives and monitoring deviations from the agreement to collude. This means that cartel members must find a way to communicate on an ongoing basis during the cartel's lifecycle. Against this backdrop, the discussion in the next section starts by examining how South African cartels communicated during their periods of collusion. This is followed by an appraisal of other structures used by the cartels, such as compensation schemes, joint ventures, sub-contracting, and other arrangements, as well as multi-market collusion. The examination also looks at how the cartels addressed cheating and the vertical features of collusion.

2.4.1.1   Communication

Communication between cartel members, whether explicit or implicit, is an essential part of collusion. It can be direct or indirect. Communication helps cartel members reach common understandings, co-ordinate the actions of cartel members, monitor defections from the agreement, and restore the collusive arrangement if cheating occurs. It is necessary to keep the collusive incentive structure intact. Evidence from South African cartels is consistent with the view that communication is central to successful collusion. The cartel record shows that cartel members communicated in one form or another, mostly in an explicit way.

Communication within the cartels that have been prosecuted in South Africa occurred through meetings, bilateral conversations by email and over the telephone, and in some cases, involved extensive exchange of information between cartel members. For example, the bread, milling, cement, scrap metal, upstream steel and wire cartels used a combination of meetings, emails

and telephone conversations and information exchange.[12] The cartels met on the sidelines of industry association meetings, at the offices of fellow cartel members, in public venues, such as hotels, and, in some instances, in holiday-type settings, such as fishing trips attended by members of the cartels.[13] These types of communication took place around the time when the collusion started, when there was a need to adjust prices, when the next tender came up or when dealing with instances of cheating.[14]

Some cartels left the role of co-ordination to specific individuals or firms, especially where the number of firms involved was large. For instance, in the wire cartel, one individual co-ordinated the cartel business of several cartel members.[15] The individual had long-standing relationships with the smaller cartel members, including involving cross-directorships in some of the smaller firms.[16] Joint ventures and the vertical nature of some of the cartels appear also to have been useful platforms for communication and co-ordination. Both joint ventures and the vertical nature of cartels are discussed in the sections that follow.

Some trade associations and other third parties also enhanced collusion. These institutions acted as platforms for more entrenched relationships between firms, as they interacted under the auspices of the industry body, which involved the smoother resolution of cartel problems. Trade associations are often set up to provide a platform for industry interaction between firms and to promote matters of common interest to the firms in the industry. For example, in policy settings, trade associations often co-ordinate and help reconcile the positions of members and play the role of advocating for certain policies on behalf of industry. This means that effective trade associations are those that become experienced and more effective at helping the industry find common ground for and achieve outcomes in favour of their collective members. Two areas where trade associations played a co-ordinating role that facilitated collusion in South Africa included acting as platforms for information sharing and as discussed below, promoting trade tariffs on behalf of members.

In some cartels, trade associations and independent third parties, which include government agencies, provided firms with more formalised and centralised mechanisms for information

---

[12] See, for example, Tribunal decisions in case numbers 10&15/CR/Mar10; 23/CR/Mar12; 37/CR/Apr08; 31/CR/May05; CR092Jan07/SA090Aug16.
[13] See, for example, Tribunal decisions in case numbers 31/CR/May05; 23/CR/Mar12; 10&15/CR/Mar10; 15/CR/Feb09.
[14] See, for example, Tribunal decision in case numbers 15/CR/Feb09; CR093Jan07/SA004Apr18.
[15] See, for example, Tribunal decision in case number CR093Jan07/SA083Aug19.
[16] See, for example, Tribunal decision in case number CR093Jan07.

sharing. These information-sharing activities enhanced monitoring by at least enabling each firm to track its own market share. The World Bank (2016) reported that 33 per cent of cartel cases between 2005 and 2015 involved the participation of a trade association. Prosecution data at the level of the firm shows that about 28 per cent of the prosecutions for collusion between April 1999 and March 2018 involved a trade association or other third party.[17] Cartels in milling, bread, steel and cement had extensive information-sharing arrangements, which included information relating to production or sales volumes, average production costs and average costs of key inputs.[18] This information exchange over time allowed firms to develop better knowledge of the costs of their rivals and provided a benchmark, or focal point, for key costs and changes over time. In the steel cartel, members shared information on capital expenditure programmes, which would have provided an indication on any capacity changes. The cement cartel even refined its information sharing arrangement to address monitoring gaps.[19] The impact of these forms of information sharing on enhancing the ability of cartels to function more effectively are readily integrated into the theory on collusion and are observed in some empirical studies.

By using customer contracts, firms can use price-matching clauses to detect cheating by other cartel members. Unsuspecting customers wishing to obtain a better price from their longstanding supplier can approach the supplier with evidence of better offers from competing suppliers, giving the supplier an opportunity to match or better the offer. In so doing, the customer inadvertently alerts to the supplier of cheating by fellow cartel members. For instance, in the wire cartel, some of the customer contracts had price-matching clauses that required customers to provide written evidence of better offers. This not only helped detect cheating using concrete proof, but also allowed the firm to respond in a targeted way rather than engage in an indiscriminate price war. It also helped avoid being fooled by customers trying to bargain through bogus claims of better offers. This shows that relationships between the sales teams of cartel members and customers can also make it easier to obtain information on the activities of cartel members, helping to detect and address cheating in a targeted way. Customer-based mechanisms can be highly effective tools for detecting and addressing cheating by cartel members, as they allow for a targeted, less costly response to cheating.

---

[17] Trade associations are more prevalent in many markets than suggested by these numbers.

[18] See, for example, Tribunal decisions on case numbers CR092Jan07/SA090Aug16; 23/CR/Mar12; 10/CR/Mar10 and 15/CR/Mar10.

[19] See, for example, Tribunal decision in case number 23/CR/Mar12.

Economic theory recognises the importance of communication in collusion: the need to convey enough information to help cartel members to reach terms of co-ordination and to monitor cheating. However, communication in most cartels is not always complete, meaning that some aspects of firms' behaviour are not monitored perfectly. For example, firms may agree to fix list prices without necessarily having the ability to monitor the discounts granted by each firm, creating possibilities for cheating to happen through discounting. This suggests that cartels that rely on monitoring list prices could experience challenges and face collapse, yet this has not been the case with most cartels.

A review of the bread, cement, milling, and steel cases referred to above showed that cartel members exchanged a large amount of information. This information exchange enabled the tracking of each firm's market share and cost information, despite the cost information being exchanged less frequently, often annually. Over a long period of time, during which firms repeatedly share information and where production processes do not dramatically differ or change, it can be expected that, even with imperfect information, firms will build up knowledge of the businesses of their rivals and of key indicative production costs. For example, in the prosecution of Pioneer Foods for participating in the bread cartel, the Tribunal observed that the costs of operating a plant bakery were comparable across different firms and were subject to the same fluctuations.[20] This meant that Pioneer Foods had a reasonable understanding of the production costs of the businesses of its rivals and the margins involved. This can be expected to enhance the ability to collude and the effectiveness of collusion.

Firms can also make use of the basic market intelligence infrastructure, through their sales teams, to track the prevailing market conditions, the discounting practices taking place in the market and to identity the firms that pursue sales more aggressively by discounting. That is, the market intelligence infrastructure can easily be directed towards collusive ends. Cartel members consider information imperfections when choosing co-ordinating or monitoring practices. For example, cartels such as the wire cartel colluded on both list prices and discounts, entailing effectively fixing the net price paid, while in other cartels (e.g. construction), the firms fixed the level of profit. Some cartels (e.g. the cement, steel, bread, and milling cartels) invested significantly in putting in place extensive monitoring mechanisms, introducing additional levels of monitoring where necessary.

---

[20] See Tribunal decision in case numbers 15/CR/Feb07 and 50/CR/May08.

It was evident in some of the cartels that communication and monitoring did not always need to take place through a centralised or institutionalised structure, such as a trade association, to help firms collude effectively. There were instances in some South African cartels where cheating was detected by the firms that were victims of cheating, rather than by a centralised cartel structure. The detected cheating could then be brought to the attention of the other cartel members, or efforts would be taken to address the situation on a bilateral level between the perpetrator and the victim (e.g. in the wire cartel). The same can be said of supplier-customer contracts that help detect cheating through price-matching clauses. Contracts are unlikely to be centralised or shared widely among cartel members. The cases reviewed in this section also show that cartels can use a combination of mechanisms to communicate and monitor the activities of members. For example, cartels that exchanged information through centralised systems (e.g. a trade association) also communicated and monitored the day-to-day behaviour of cartel members through various means, which included the use of sales teams, price-matching clauses, and bilateral communication.[21] Economic theory does not always explicitly incorporate these complementary elements of communication and monitoring that a cartel could use in some combination, depending on its circumstances. This could explain the limited extent to which economic theory can explain collusion where some aspects of firm behaviour are not readily observable through centralised structures of communication and monitoring.

Economic theory[22] predicts that collusion is exceedingly difficult to sustain in the face of uncertainty. This theory places more weight on perceived uncertainties where certain aspects of firm behaviour are not readily observed. This is especially the case where the collusive mechanism leaves room for rivals to cheat or where the formal monitoring mechanism is imperfect, making the detection of cheating challenging. Evidence from South African cartels suggest that cartels mitigate the uncertainties using other complementary, non-centralised mechanisms such as market and business knowledge built over time; firm-level mechanisms such as customer relationships; and basic firm-level market intelligence infrastructure, which firms that compete ordinarily possess.

South African cartels used different forms of communication and monitoring, and in some key cartels, different forms of communication and monitoring complemented each other to provide the cartels with a more complete structure for sustaining collusion, even in situations where

---

[21] See, for example, Tribunal decision in case number 15/CR/Feb07 and 50/CR/May08.
[22] Levenstein and Suslow (2011) point out that *"Economic theory identifies uncertainty as the primary cause of cartel instability (Stigler 1964; Green and Porter 1984; Abreu, Pearce, and Stacchetti 1986)."*

some aspects of firm behaviour were not readily observable to the cartel members. For example, in a price-fixing cartel monitoring through exchanging sales volumes, sales teams through their interactions with customers provided useful information on the discounting practices of cartel members. This complements the use of sales volumes as a monitoring tool.

2.4.1.2 Compensation, or payment schemes; joint ventures; sub-contracting; and other arrangements

South African cartels have adopted different structures for aligning incentives and keeping them as such. Compensation or payment mechanisms; joint ventures; and sub-contracting arrangements were features of some key cartels. A compensation scheme can be defined as a tool used by cartels to disincentivise firms from cheating, by transferring the gains from cheating back to the firms that are victims of the cheating [see, for example, Porter (2005), Levenstein and Suslow (2011)]. The goal of this is to maintain the agreed apportionment of collusive rents, which in turn helps keep the incentives to collude intact.

In this study, compensation or payment schemes were defined more broadly than the narrow transfers to other cartel members by a cartel member who either cheats or finds themselves with greater rents, due to demand fluctuations. The author included those situations involving payments to fellow cartel members in the form of a share of profits from contracts and payment of a loser's fees to firms that agreed to provide cover-prices in tenders. The use of such schemes ensures that collusive rents are shared in the agreed form among cartel members and, in doing so, helps to keep the incentives to collude compatible. Examples of the use of such payment schemes include those in the construction sector, as highlighted by Ratshisusu (2014).

Collusion in construction contracts, for example, included an agreement for cartel members to provide cover-prices and, on top of that, ensure that the firm that agreed to cover the price gets a share of the profit, which in at least one instance was achieved through sub-contracting some of the project's work to the firm that provided a cover-price.[23] It appears that in this way, the cartel solves several problems. The first problem solved by the cartel is that it avoids price competition through cover pricing, enabling the cartel to realise the price it wants to charge. The second problem that the cartel solves relates to the need to find a mechanism for sharing collusive rents among participants. Third, by sub-contracting the cartelised work and sharing profits with the firm that provided cover-prices, the cartel also resolves a trust issue among its

---

[23] See, for instance, the Tribunal decision in case number 016949.

members. The cartelisation of some construction contracts involved paying losers' fees to the firms that provided cover-bids.[24] The former CEO of one of the firms testified about a cartel scorecard that was used for tracking outstanding payments, which were either invoiced or offset against other projects [see Ratshisusu (2014)].

While Ratshisusu (2014) saw the use of sub-contracting and payment of losers' fees as a form of collusion, in this dissertation the author argues that these arrangements are better characterised as part of the management of the collusive structure – ways of aligning incentives and sharing collusive rents. This would be more consistent with the theoretical and empirical literature on cartels using payment schemes to keep incentives compatible. Such an approach also helps to distinguish between forms of collusion (e.g. price fixing, market allocation and collusive tendering) and management structures, which are used to sustain collusion.

Joint ventures facilitated collusion in several ways. First, they reduced the number of firms independently bidding for contracts or independently competing in markets. Joint ventures also replaced the independent actions of competing firms with co-ordinated, dependent actions. Second, they helped to co-ordinate the interests of various firms that, through joint ventures, aligned their interests, especially in the construction sector. Third, joint ventures and cross-directorships created avenues for communication and information sharing between firms. Through joint ventures, communication between firms can be direct and informal, yet more effective, given that the interests of the firms are aligned through the joint venture. Fourth, collusion through joint ventures can be difficult to detect given that they can serve both a legitimate pro-competitive commercial purpose and an illegitimate collusive purpose. The two purposes can be difficult for competition authorities to distinguish. Finally, joint ventures created easier avenues for generating and sharing collusive rents in a more transparent manner, since each joint venture partner would get an agreed proportionate share of the rents. The prevalence of joint ventures between competing firms in construction projects suggests that joint venture arrangements served as a mechanism for sharing rents in selected contracts, and in some instances, cartel members outside the joint venture got a share of the work through sub-contracting arrangements.

---

[24] See, for instance, the Tribunal decision in case number 017038; 016949.

Collusion in soda ash, poultry breeding, print media, waste removal, steel, ocean transport, construction and cement involved joint ventures.[25] Notable in the construction sector was the routine way in which the major construction firms participated in joint ventures that were used for collusion. This suggests that joint ventures provided a simple and effective way of reconciling the interests of firms, sharing collusive rents, and accommodating the interests of various firms. In the steel-to-wire value chain, an upstream cartel operated with two members of the cartel having a joint venture that operated in a downstream cartel. The downstream cartel also included the downstream activities of another member of the upstream cartel. The vertical nature of collusion is discussed in a section that follows.

Some collusive arrangements, while not fitting perfectly within the framework of joint ventures or sub-contracting arrangements, appear to have had similar effects as mechanisms for aligning the incentives of firms and for distributing collusive rents. These included the use of toll-manufacturing arrangements in collusive conduct between Sasol and Foskor, and the use of supply arrangements that involved supplying products to, or through other cartel members, as was the case in the fertiliser, poultry breeding and wire cartels.

The use of joint ventures as part of collusive arrangements calls for greater scrutiny of joint ventures by competition authorities. Part of the scrutiny should involve proper characterisation of joint ventures to determine whether they are in place for legitimate purposes or for anti-competitive purposes. In doing this, the competition authorities should focus on characterising the nature of the conduct given the economic relationship between the firms, the potential effects of the conduct on the competitive process and the likely economic effects on market outcomes. This assessment should be undertaken in the context of a critical examination of the rationale for the existence of such arrangements as well as any restraints that may be contained in such arrangements. Such an approach ensures that the potential impact of joint ventures in aiding collusion is grounded in economics theory behind the prohibition of collusion and that only those joint venture arrangements that are likely to lead to collusive outcomes are characterised as such. The same approach can be applied to the use of other arrangements such as toll-manufacturing, sub-contracting and supply arrangements between competitors. It is important to note that the conclusions that can be drawn from the application of the suggested

---

[25] See for example, the Tribunal decisions in case numbers CO044Jun13; 017277/CO077Jul13; 23/CR/Mar12; 016949; 49/CR/Apr00; the wire cartel; the cement cartel, etc.

framework will depend on the facts of the case. It avoids a blanket approach to arrangements between firms which may lead to both Type I and Type II enforcement errors.

The foregoing discussion is consistent with the hypothesis that the use of compensation or payment schemes, joint ventures, sub-contracting, and toll-manufacturing and cross-supply arrangements can facilitate collusion. They help align incentives and share collusive rents. The mechanics and definition of schemes that involve payment or compensation may vary from case to case. In some cases, compensation schemes could simply refer to arrangements where a cartel member that cheats or realises greater sales than its agreed share transfers a portion of its rents to other cartel members. A review of the cartel record in South Africa suggests that the notion of a payment or compensation scheme could be broader to include compensation for foregoing a contract and entitlement to a share of profits from the cartelised projects. One common feature of these arrangements is that they keep the incentives to collude aligned and facilitate sharing collusive rents in an agreed way. This would be consistent with the rationale for the use of compensation schemes in Levenstein and Suslow (2006,2011) and Porter (2005).

2.4.1.3 Multi-market and multi-product[26] nature of cartels

Firms that have multi-market contact interact repeatedly over time and across markets. A significant number of cartels involved firms that enjoyed multi-market contact, as well as multi-product firms in which multiple products were cartelised. Examples of cartels that involved multi-market contact include those in the steel industry, pharmaceutical industry, agriculture industry, manufacturing industry and construction industry. This is consistent with the theoretical and empirical literature, which shows that collusion is more likely to be possible to sustain in the presence of multi-market contact.

As discussed in section 2.2, multi-market contact creates better opportunities for aligning the incentives to collude and for more flexible ways of addressing cartel problems, including the distribution of collusive rents, and dealing with cheating. For example, in the fertiliser cartel, the collusion resulted in Sasol becoming the sole supplier of limestone ammonium nitrate (LAN (28)) to the wholesale market.[27] The firms involved in upstream steel cartels colluded in

---

[26] There is notionally a distinction between multi-market collusion and multi-product collusion, in that products in a multi-product cartel may or may not be in the same relevant market. If the products are in separate relevant markets, this would be characterised as multi-market collusion. If they are in the same relevant market, this would just be multi-product collusion.

[27] See Tribunal decision in case number CR006May05/SA095May09, which included case number 31/CR/May05.

the production and supply of inputs to the steel-to-wire value chain as well as to the steel-to-reinforcing bar value chain. The upstream steel cartel also had market allocation features. The largest millers of maize and wheat products were also involved in collusion at the downstream level, where they also competed in the production and supply of retail products such as bread and mealie meal. A review of collusion in the construction sector best illustrates this because collusion across different types of construction activities allowed firms to allocate contracts across different activities through reciprocal cover-pricing and over time.

As stated in section 2.3, 42 per cent of firms prosecuted between April 1999 and March 2018 engaged in serial collusion or were part of a group in which another firm participated in collusion in another market. The World Bank (2016) reports that for cartels prosecuted in South Africa between 2005 and 2015, in 63 per cent of the cases, a firm was also in another cartel in a different market of the same sector, or in another sector. The World Bank (2016) uses the network mapping technique based on firm ownership links to show the links between colluding firms in the agriculture and food sectors. Such analysis reflects the extent of multi-market contact between firms that may be colluding in the same cartels, and it shows how multi-market contact can lead to the export of collusion as a rational business strategy to other markets or sectors.

The South African cartel record is consistent with the theoretical and empirical literature, which shows that cartel conduct is possible to sustain where firms have multi-product/-market contact. This is particularly so for the cases reviewed in this section, which were mostly characterised by multi-product firms that competed across multiple markets.

2.4.1.4   Cheating

As discussed in section 2.2, economic theory shows that cheating can be less profitable than honouring the collusive agreement. Cheating is not a feature of cartels that is commonly reported in South African case decisions, potentially supporting theoretical predictions. However, this may be more a result of the way most prosecutions are finalised rather than evidence of the absence of cheating. Cases that are concluded through settlement seldom explore the detailed nature of a cartel's problems and how it resolved those problems. However, if cheating had posed a significant risk to collusion, one would have expected that the firms involved would have raised the disruptive nature of cheating as part of the negotiations or arguments for lower penalties, as was the case with the wire cartel.

39

The cement and wire cartels are two cases where cheating was a significant problem, that it was discussed in some detail in the cases. The discussion of cheating in the cement cartel arose because the solution to such cheating was an adaptation to the collusive arrangement and its monitoring mechanisms, which coincided with the transition from a legal cartel to an illegal cartel. This also coincided with the beginning of the modern era of cartel enforcement in South Africa.

Cheating in the wire cartel was debated following arguments raised by firms that were facing prosecution. The firms argued that the price war that had resulted from cheating signalled a breakdown of the collusive arrangements, as opposed to being a means of managing the collusive arrangement. Cheating in the wire cartel appeared to have been triggered by a slump in demand for at least one of the cartel members, who became financially impatient and decided to deviate from the collusive arrangement. In both cases, collusion involved multiple products, and the price wars involved more than one product.

Mondliwa and Das Nair (2019) reported that the reinforcing-bar cartel resolved instances of cheating by targeting the customers of the firm that deviated from the agreement to collude. Such retaliatory mechanisms continued until the firms regrouped to restore the collusive agreement. These cases show that price wars were useful tools for keeping or restoring the compatibility of the incentives to collude. In the cement, wire and reinforcing-bar cases, cheating and the ensuing reprisals did not end the collusion but were succeeded by adapted collusion in the case of the cement cartel and a reversion to the same collusive scheme in the cases of the wire and steel cartels. The three cartels were ended by the intervention of the competition authorities.

Some of the cartels that engaged in price fixing adopted simplified pricing structures. In industries such as the steel and steel-related markets, pricing was based on simple import-parity pricing and delivered-price rules, with changes to prices following changes in international steel prices. A standard-pricing formula was used by the scrap metal cartel to fix the purchase price of scrap metal. Price changes that followed changes in international steel prices not only helped to simplify the pricing structure but also helped cartels to avoid detection, since the changes followed a historically industry-accepted notional cost of steel, which customers also accepted. The cement cartel used base-point pricing rules. In a cartel like the wire cartel, a simple framework of agreed list prices less agreed discounts was applied to all cartelised products, with price changes mostly being informed by changes in the international price of

steel products. Simplified structures, coupled with information exchange, involving repeated interaction across markets and over time, can mitigate the detrimental effect of cheating on a cartel. For some cartels, meetings and telephone discussions were sufficient to resolve cheating problems. For others, resolving cheating problems required adapting arrangements, including investing in greater information sharing.

Cheating occurs in cartels, but it does not always lead to the death of a cartel [see, for example Levenstein and Suslow (2011)]. Cartels find ways of addressing cheating. This can be through price wars, or enhanced monitoring capabilities, or compensation schemes. Some of the cartels considered in this chapter experienced cheating, but this did not lead to the death of the cartels. Cheating was addressed through price wars and enhanced monitoring capabilities. Other factors could also have mitigated the temptation to cheat. These included the prevalence of multi-market contact, joint ventures, payment schemes and the vertical nature of collusion.

Cheating has not been widely reported in South African cartel cases, except for a few. In the cement and wire cartel cases, it appears that the price wars were symmetrical, involving all the cartel members moving to a non-collusive phase, at least with some products. In the rebar case, the cartel members targeted the customers of the cheating member. In all the cases, the phases of cheating and retaliation did not cause cartel death but re-established the agreement. The observation that cheating is not the pre-eminent cause of cartel collapse was also observed in studies such as Levenstein and Suslow (2006 and 2011).

2.4.1.5 Vertical features of some South African cartels

A feature of some of the key South African cartels was the vertical nature of collusion. Several of the major cartels involved cartel members that were vertically integrated and were involved in cartels either at both levels of the value chain or where a firm was dominant upstream was involved, in some way,[28] in collusion at the downstream level. Examples of cartels in which the same firms or some of the firms were involved in cartels both downstream and upstream included those in the steel and wire industries; cartels in agriculture and the agro-processing industry; cartels in maize and wheat milling, which were also involved in bread and maize milled products; the wooden products cartel, in which an upstream cartel supplied raw board to downstream competitors that they also competed against; and cartels in the chemical-to-

---

[28] This could be directly through itself or a wholly-owned subsidiary, or indirectly through, for example, some joint venture activity.

fertiliser value chain. The major construction firms were also vertically integrated, with involvement in collusion at various levels of the value chain. The upstream cartel or major upstream firms supplied key inputs to the downstream cartel, which the firms were also part of. The glass cartel involved a dominant upstream firm that was also involved in or appeared to support downstream collusion. This phenomenon raises economics questions, explored below.

Upstream and downstream markets are complementary in nature, and vertically integrated firms are, in theory, expected to have the incentive to reduce prices [Bishop and Walker (2010)]. This is because a higher price at the upstream level reduces demand for the product upstream and reduces demand for the downstream product. With this economic effect, both the upstream and downstream businesses benefit, through higher sales volumes and revenue, from a lower price being charged at either level of the value chain. A cartel at each level of the value chain is expected to raise prices above the competitive level, meaning that the existence of each cartel undermines the demand for its own product and the product at another level. Given this effect, the incentives to participate in cartels operating at complementary levels of the value chain are not obvious.

The firms that colluded at upstream and downstream levels were significant suppliers of the respective products at each level and supplied inputs to other downstream firms. For example, firms that operated the upstream wheat milling cartel were the largest suppliers of flour products, and the same firms also supplied a significant share of supply of bread products (estimated to be between 50 per cent and 60 per cent of the domestic bread market).[29] They also supplied flour to some downstream independent bakeries, which they competed with. A similar observation can be made for the upstream steel cartels, which also competed in downstream markets, where they were major suppliers of steel input products. Sasol and its competitors supplied inputs to their downstream competitor, Nutri-Flo.

The supply of key inputs to downstream competitors gave upstream firms that also participated in downstream cartels influence over the costs of the key inputs of downstream members of the cartels. Control over the supply and costs of key inputs potentially helps address incentives to cheat by smaller non-vertically integrated rivals and to marginalise those that attempt to undermine the cartels. It also means that the smaller cartel members potentially become

---

[29] The remainder of the market was supplied by about 4,000 smaller independent bakeries comprising stand-alone, in-store and franchise bakeries (see the Tribunal decisions in case numbers 10/CR/Mar10 and 15/CR/Mar10).

followers of the larger vertically integrated cartel members. In a situation like the bread cartel, independent bakeries faced marginalisation in two ways. First, through upstream cartel overcharges, which Mncube (2014) estimated to range between seven per cent and 42 per cent and, second, through exclusionary behaviour.[30] This mitigates the risks that firm asymmetries pose to collusion, as smaller firms may find it better to align with the cartel arrangement than attempt to be mavericks. The vertically integrated structure of collusion, where the larger firms control most of the supply, means that these firms have greater control over value chains and distribution networks, rather than merely being content with cartelising only one level of the value chain. This enhances the probability of the success of collusion.

The use of supply and production agreements also appeared in cartel arrangements. Some firms in the wire cartel used distributorship arrangements, where one cartel member would exit some geographic regions and have their products distributed exclusively by a fellow cartel member in those areas. The firms saw exclusive distributorship as a solution to incidents of price competition (cheating). Firms in the fertiliser and related-chemicals cartel used toll-manufacturing and supply agreements. While these arrangements may appear to be vertical in nature, they are, in effect, horizontal co-ordination arrangements reinforcing collusion.

Participation in collusion at two successive levels of the value chain weakens or eliminates competition at both levels. This may result in higher prices overall along the value chain, making vertically integrated cartels more successful. Operating at two levels means that the cartel can capture upstream sales that are lost when independent downstream firms are foreclosed. The ability of effective competition at one level to potentially undermine collusion and the exercising of market power at another level is weakened, as the same firms control production and supply at both levels. In this study, the author concludes that the observed vertically integrated cartel structure of some cartels enabled the cartels to exercise market power better than would have been the case if the firms had colluded at only one level.

The vertical features of collusion (firms colluding in cartels operating at successive levels of the value chain) are not widely discussed in economic theory and in empirical studies. Discussions about the impact of vertical integration and restraints on collusion mostly focus on helping firms become more symmetrical, enhancing transparency through information sharing and raising entry barriers. These would appear to remain relevant in the case of the cartels

---

[30] See, for example, the Tribunal decisions in case numbers 10/CR/Mar10 and 15/CR/Mar10.

discussed in this section, but they do not necessarily address the incentives of firms to collude at successive levels of the value chain. The inference drawn in this study from collusion at successive levels of the value chain is that these institutional features enhanced the success of firms at colluding.

2.4.2 Dealing with threats of entry

Cartels try to charge supra-competitive prices and earn higher margins. If a cartel lasts for several years, higher margins are expected to attract entry by firms outside the market. Expansion by those outside the collusive arrangement in turn destabilises the cartel. Levenstein and Suslow (2006) found entry to be a significant contributor to the death of cartels. Effective cartels succeed in restricting the entry and expansion of firms outside the collusive arrangements. The World Bank (2016) reports that most cartelised markets in South Africa were highly concentrated, with high entry barriers. The focus of this dissertation was on examining how South African cartels deployed strategic entry barriers over and above other forms of entry barriers that characterised the markets. From the cartel data and developments in trade tariffs, the author identified two ways through which cartels could have deployed strategic entry barriers. First, they could have done so in the form of exclusionary conduct and, second, by advocating for, or supporting protectionist instruments such as trade tariffs.

2.4.2.1 Exclusionary behaviour

The evidence from some cartels shows that cartels used a variety of entry and expansion deterrence strategies. These strategies can broadly be categorised into predatory behaviour, price discrimination and margin-squeeze types of exclusionary conduct. The latter types of exclusionary conduct were facilitated by the vertically integrated nature of the leading firms in some cartels. Some of these firms also supplied key inputs to their downstream competitors.

In the bread cartel, one of the larger firms, Pioneer Foods, employed predatory strategies against smaller independent bakeries, marginalising them in the process.[31] Part of the allegations in the bread cartel was that the agreement also included not supplying bread to new bread distributors and to one another's former employees. This conduct appeared to be aimed at limiting the disruptive impact of new distributors and former employees on collusion.

---

[31] See the Tribunal decisions in case numbers 10/CR/Mar10 and 15/CR/Mar10.

Although it was not prosecuted, Arcelor Mittal was accused of engaging in exclusionary conduct by one of the smaller maverick downstream competitors. Notably, in its settlement with the Commission, Arcelor Mittal agreed to remedies relating to its pricing conduct.[32] It must be stated that the complainant reserved their right to review the settlement agreement between the Commission and Arcelor Mittal.[33] Similarly, in the wooden board industry, two duopoly cartelists[34] in the supply of raw wooden board, who also competed in the downstream market, were accused of engaging in exclusionary conduct through margin squeeze-type conduct.

In the fertiliser value chain, Nutri-Flo complained that it was subjected to exclusionary behaviour through excessive pricing conduct linked to the pricing of inputs.[35] This conduct appeared to have been supported by the vertical nature of collusion, where some cartel members were suppliers of critical inputs to their own downstream competitors, who in turn, were either in the collusive ring or outside of the collusive ring. The vertical nature of collusion in this case helped with managing the internal dynamics of the collusive arrangements, as discussed above, and with limiting the impact of external threats to the cartels.

Another feature of entry or expansion deterrence is the use of vertical restraints to deny new entrants access to customers or suppliers. The scrap metal cartel used exclusive-supply agreements to limit the supply of scrap metal to customers outside the collusive arrangement and to restrict the purchase of scrap metal from suppliers outside of the collusive ring. In markets that are characterised by economies of scale and scope, this conduct can effectively neutralise the threats that entry and expansion can pose to a cartel.

2.4.2.2 Strategic use of protectionist instruments such as trade tariffs

In some South African markets (e.g. steel, grain, textiles, etc.), imports present the next best alternative for customers, and historically, pricing was import-parity-based. For some cartels, the threat of import constraints was limited by import tariffs. A review of the post-1999 era of regulation shows that firms in some cartelised markets, either directly or through third parties such as trade associations, opposed the removal or reduction of import tariffs and in some

---

[32] See the Tribunal's press release dated 16 November 2016, available at: https://www.comptrib.co.za/case-detail/7177.
[33] Ibid.
[34] See the Tribunal decision in case number CO030May19.
[35] See the Tribunal decision in case number 31/CR/May05.

instances applied for the increasing of tariffs. In this instance, the behaviour appears to have continued even after the cartels had ended, suggesting that firms may have had the incentive to limit competition from imports even long after the cartel conduct had ended.

During the period of collusion, tyre manufacturers through their industry body (SATMC), opposed a reduction in tariffs on some tyres.[36] The same manufacturers even opposed the reduction of tariffs on imported second-hand types, partly because of the rising incidence of second-hand tyres and the downward impact on the price of new tyres.[37] Similarly, during the period of collusion, the Association of Electric Cable Manufacturers of South Africa opposed an application for rebates on tariffs on cables and wire products classifiable under the broader tariff heading, which could have been read to include imported products that competed with their products. The association argued that the rebates would have increased imports and "caused injury" to the domestic industry.[38] Even in the post-cartel era, some of these firms continued to request tariff increases or to oppose the reduction or removal of tariffs, or the introduction of rebates on imported competing products. For example, in 2010, around the time the electric cables cartel was reported to have ended, the Association of Electronic Cable Manufacturers of South Africa also raised concerns that the introduction of import rebates on insulated electric cables would, among other considerations, have eroded support provided to domestic manufacturers of cables.[39]

In 2005, the International Trade Administration Commission of South Africa (ITAC) undertook a review of the customs tariff regime that applied to wheat, wheat flour and wheat-related downstream products. The review task team included industry participants and bodies such as the National Agricultural Marketing Council (NAMC). The review showed that wheaten flour and wheat-related downstream products enjoyed significant tariff support. Although the proposal was rejected by ITAC, the industry advocated for the raising of import tariffs on wheat and wheat flour products.[40] At the time, South Africa was a net importer of wheat, meaning that higher wheat prices would have negatively affected wheat milling

[36] http://www.itac.org.za/upload/document_files/20141015044013_Report-84.pdf.

[37] http://www.itac.org.za/upload/document_files/20141015051458_Report-99.pdf.

[38] http://www.itac.org.za/upload/document_files/20141013065604_Report-106.pdf.

[39] http://www.itac.org.za/upload/document_files/20140928120142_Report-349.pdf.

[40] http://www.itac.org.za/upload/document_files/20141013063334_Report-112.pdf.

businesses.[41] It is notable, however, that wheat-milling businesses benefited from the rebates applicable to wheat, which lowered their costs of wheat. ITAC noted that imports were not posing significant competition to the wheat milling industry, which enjoyed tariff protection. The wheat milling industry continued to enjoy significant tariff support after the review.

ITAC undertook a review of tariffs in the steel industry in 2005, observing that the industry, which relied on import parity pricing, was protected from international competition through the geographic isolation of South Africa and a five per cent tariff that applied to imports, which it regarded as superfluous.[42] The review sought to find ways of reducing the cost of steel that downstream industries faced and considered removing the tariffs on imported primary steel products. SAISI, the industry body, opposed removal of the tariffs, arguing instead for the maintenance of the tariff. Columbus Stainless and the Southern African Stainless Steel Development Association opposed removal of tariffs on stainless steel products, cautioning that this could negatively affect the viability of primary producers.

These observations do not undermine the policy use of tariffs to protect the legitimate interests of any nation and the right of firms to fight for issues of interest to the markets they operate in. It calls for close examination of the legitimacy of the interests represented in these actions – that is, in opposing the reduction or removal of import tariffs or supporting the imposition or increase of current tariffs. Tariff support is one way that firms may rely on to dampen competition, especially in those markets where domestic entry is less likely to occur. For example, in 2014 the South African Wire Association and some members of the erstwhile wire cartel supported an application by one of its members to introduce tariffs on imported wire products.[43] Similarly, in 2015 SAISI on behalf of the upstream steel producers who were part of the upstream steel cartel, together with Highveld Steel and Vanadium Limited, applied for tariffs on various products.[44]

Tariff protection provides a costless mechanism for cartels to manage the external threats posed by imports of inputs and final products that compete with the products of cartel members. The

---

[41] Premier Foods commented about the upward cost impact of such moves; the concern appears to have been more about the input cost impact on its wheat milling business, and not necessarily the higher prices that consumers would have faced.

[42] http://www.itac.org.za/upload/document_files/20141013061359_Report-121.pdf.

[43] http://www.itac.org.za/upload/document_files/20150218092136_Report-No-487-Increase-in-the-rate-of-customs-duty-on-certain-wire-products.pdf.

[44] http://www.itac.org.za/upload/document_files/20160108084221_Report-No-509.pdf.

cartel does not need to engage in any exclusionary conduct, which results in a sacrifice in short-term profits, yet it could still effectively partially support collusion.

2.4.3 Summary of key lessons from the characteristics of detected cartels and institutional features of key cartels

Several lessons can be drawn from the review of the characteristics of detected cartels and the institutional features of collusion. The features of cartels include the following:

(i) Cartels employed a mixture of collusive practices, which were probably driven by the nature of the transactions with customers and the ease with which the co-ordinating mechanism could be used.

(ii) Collusion happened across all sectors of the economy, but was concentrated in manufacturing, trade, transport and construction. The communications and health sectors had the lowest cartel detection rates. The number of firms per cartel varied widely, from two to as many as 20 firms of differing sizes. The large number of firms in some cartels might be because show that for the cartels to be able to control market prices, they needed to incorporate as many firms as possible. However, in some sectors, this might simply have reflected the continuance of a culture of co-operation that predated the current competition law regime.

(iii) On average, most cartels lasted several years, with average firm participation in a cartel between 1999 and 2018 being six years, but this number varied by cartel, firm, and sector. The shortest cartel lifespan was less than a year, and the longest lifespan was 17 years.

(iv) Evidence from studies of cartel overcharges in South Africa shows that cartels were able to exercise market power to significant but varying degrees.

The institutional features of detected cartels and selected key cartels reveal several lessons, which are as follows:

(i) Communication in detected South African cartels took place using meetings, telephonic discussions, emails and, in some cases, involved extensive information-sharing structures. Communication was not always centralised and could have been bilateral and informal. Cartels can use a combination of different forms of communication, for example, meetings, telephone discussions and extensive information-sharing arrangements. Customers can also be a rich source of information relating to the

conduct of fellow cartel members who try to cheat on collusive arrangements. The normal sales team infrastructures and relationships can help cartels monitor and detect cheating by fellow cartel members, and this monitoring can happen at a firm level, and not necessarily at a centralised, cartel level. As such, cartels do not necessarily need to invest in costly, extensive and centralised monitoring systems.

The theoretical literature addressing collusion under imperfection should be expanded to consider how using combined mechanisms of communication and monitoring could succeed in closing the gaps left by incomplete information, making collusion more successful. For example, firms that collude on parts of a price (e.g. list prices) could monitor the uncertainties of discounting practices using their sales team infrastructures and customer contracts with matching clauses to complement monitoring through an exchange of sales information. Cartels are not restricted to using just one centralised mechanism. Similarly, firms that collude repeatedly over time could build sufficient knowledge of the businesses and cost structures of their rivals, especially if production processes are similar and do not change significantly over time. Further, there is need to examine the effectiveness of indirect communication in facilitating collusion. This can be done by examining the extent of overcharges in those cases where indirect communication is employed compared to those cases where cartel members have more direct communication, controlling for other factors that impact on overcharges.

(ii) Some cartels use compensation/payment schemes, joint ventures, sub-contracting, and other arrangements. These are similar to those shown in the literature on other types of arrangements [e.g. Levenstein and Suslow (2006,2011) and Porter (2005)]. These arrangements are useful mechanisms for keeping incentives compatible and distributing collusive rents among cartel members.

(iii) Multi-market or multi-product collusion is a feature of many cartels. This is consistent with the theoretical and empirical literature. There is a high likelihood that firms that interact in several markets will collude in other markets after these firms have formed a cartel in one market.

(iv) Cheating happens in cartels, but it is not widely reported as a big problem that threatened and led to the death of detected cartels in South Africa. In cases where it was reported, it did not lead to the death of the cartels. Ensuing price wars led to the re-establishment of the collusive agreement, rather than the death of the cartels.

(v) Some key detected cartels exhibited vertical features of collusion, where firms collude at successive levels of the value chain. Cartels seek to raise prices above competitive levels, and this is expected to have the effect of reducing the demand for the cartelised products and those of the complementary products at the next level of the value chain. This means that firms operating at each level of the value chain are expected not to condone or participate in a cartel at the next level of the value chain. The participation of the same firms in cartels at multiple, successive levels of the value chain suggests that relaxing competition at successive levels in some instances enhances the prospects of success or the effectiveness of collusion, with collusive rents being distributed between the upstream and downstream levels of the value chain; otherwise, this feature would not be present. This feature of collusion is not widely addressed in the economics literature.

(vi) As expected in the economics literature, cartels deal with external threats through exclusionary conduct, which may be supported by the vertically integrated nature of larger firms involved in collusion that also supply key inputs to their rivals at the next level of the value chain. In South Africa, imports offered the next best alternative source of supply for inputs, and import tariffs appear to provide a way of mitigating the impact of imports on cartel stability. A review of the activity of firms or their trade associations in the realm of trade tariffs suggests that cartels potentially saw this as a viable, costless tool for mitigating the impact of imports. The strategic use of trade tariffs to limit the impact of competition may continue even long after the cartels end.

2.5 Conclusions and policy implications

In many ways, South African cartels have been supported by a combination of many of the factors that are identified in the theoretical and empirical literature. South African cartels have also adopted institutional features that are consistent with the theoretical and empirical literature. The South African cartel record confirms that cartels are not and will never be perfect organisations. They will face various challenges, whether they be challenges in aligning disparate interests, cheating or the threat of entry or expansion by firms outside the cartels. This, however, does not make them ineffective or inherently unstable to the extent that they would cease to exist. Instead, the record of detected cartels shows that a lot of the cartels lasted several years, and case-specific evidence on overcharges shows that cartels can exercise the

collective market power afforded by collusion. Cartels find ways of managing the threats that confront them.

Of the various institutional features available to cartels in South Africa, it is worth emphasising three that were observable from the key cartels: communication, the vertical nature of collusion, and the potential use of trade policy tools by cartels.

Relating to the first of these, communication in cartels is important and was prevalent across all detected cartels. Communication takes various forms, and not every aspect of communication needs to be centralised within the cartel structure. It happens at different levels within firms and can be complemented by the various mechanisms available to firms in the ordinary course of business (e.g. sales teams infrastructure, customer relationships and learning effects built over time). Given the prevalence of cartels, the various forms that communication can take, and the various mechanisms that cartels can use to complement imperfect information, there is the possibility that economics theory presently overstates the uncertainties and complexities of collusion and consequently sets very restrictive requirements for successful collusion.

Second among the three emphasised features of the key South African cartels is that they had a vertical dimension, which appeared to contribute significantly to their successful collusion in two ways. On the one hand, it helped with aligning the internal interests of the cartel members, by exposing the smaller cartel members to the discipline of their larger fellow cartel members, who by virtue of vertical integration, controlled key inputs while also colluding in upstream markets. On the other hand, control over key inputs by vertically integrated upstream cartels placed those firms in the upstream market in a better position to deal with the external threats of entry or expansion by firms outside the downstream cartels, since the cartels essentially controlled the value chains and distribution networks associated with their products. By relaxing the competition at successive levels of the value chain, the vertically integrated cartels were better placed to succeed and exert themselves.

Third among the emphasised features of the key South African cartels was the use of trade tariffs to dampen the threats posed by imports. For most downstream rivals, the next best alternative to getting inputs from vertically integrated cartel members was imports. In this area, the input markets were in some instances protected, using import-parity pricing and trade tariffs instituted at the request of, or with the support of cartel members or the trade associations

representing the cartel members. Higher prices for imports would have made them an unattractive alternative source of supply, forcing downstream rivals outside the collusive arrangement to rely on vertically integrated cartel members for supply. This weakened the ability of these rivals to threaten the cartels. Similarly, weakening constraints from imports of final products in downstream markets weakens the threat to downstream collusion.

There are several policy implications that emerge from the examination of the institutional features of collusion in South Africa:

(i)   In cartel investigations, there is need for the competition authorities to examine the nature of communication between the employees of various firms at different levels. This communication need not be limited to physical meetings, email, and telephonic contact, but could include informal and indirect communication. Such an examination should extend to the institutional structure of trade associations and third-party bodies that suspected cartelists interact with. These third-party bodies could even include government bodies. It is important to note that information exchange does not need to be centralised within a cartel structure, or to eliminate all uncertainty to be effective in aiding collusion. Some aspects of monitoring can occur at the firm level. This is particularly the case where there are long-standing horizontal and vertical relationships between firms, where firms have sufficient knowledge of rivals' businesses, and cartel members have long-standing relationships with customers.

(ii) The second policy implication, given the vertical nature of collusion in some of the cartels, is that once there are suspicions of collusion at one level of the value chain, it is important for the competition authorities to examine the possibility of collusion at other levels, or in other markets where the same firm interacts with its fellow cartel members. Mapping cross-shareholding or cross-directorships should be undertaken, and where collusion is proven, it should be addressed with remedies. The vertical nature of collusion also places vertically integrated cartel members that supply key inputs to non-cartel members in a better position to restrict the entry or growth of a competitive fringe. This means that allegations of exclusionary conduct in markets suspected of collusion or where collusion had previously been proven should be examined, not just as standalone abuse conduct, but also as part of a coherent theory of collusion. Upstream firms may not individually meet the legal dominance threshold for abuse but could nonetheless be acting in concert to foreclose downstream rivals who threaten collusion.

52

The same applies to the use of vertical restraints between firms suspected of collusion, and between firms suspected of collusion and their customers.

(iii) The third policy implication is that cartel investigations or screening for collusion should also involve a search for the existence of joint ventures, the incidence of sub-contracting, cross-payments, cross-supply, and toll-manufacturing arrangements between firms suspected of collusion. In so doing, the commercial rationale for such arrangements should be carefully examined, especially where there are no capacity constraints. Sub-contracting arrangements should be cross-checked against bidding behaviour and tracked over several contracts that the firms have or could have participated in, rather than just the individual contracts under scrutiny. This should help build a more complete picture of the organisation and management of collusion.

(iv) The fourth policy implication is that multi-product firms that are suspected of colluding in one market should be screened for collusion in other markets, and links between collusion in various markets should be examined. This is because the management of collusion could involve a *quid pro quo* structure that spans more than one market or product and treating collusion in each market separately is likely to lead to a partial understanding of the framework of collusion employed by the firms. A partial understanding of collusion undermines deterrence.

(v) The final policy implication is that in markets where imports present the next best alternative source of supply for key inputs, other than the firms suspected of cartel conduct, investigations of collusion should incorporate an examination of the activities of cartel members or their trade bodies in advocating for or supporting the imposition of trade tariffs on competing imported products. This could be a useful and costless way for cartels to raise entry barriers. Where government is contemplating imposing tariff protection for domestic firms, it is necessary to tie such support to clear growth and developmental goals that include preserving domestic competition. Failure to do so raises the risk of supporting the stability of collusion, by protecting cartel members from the destabilising effects of import competition. The willingness and role of government in making collusion possible, whether inadvertently or as a result of balancing other policy considerations such as supporting domestic industry, requires further study. This further study should balance the anticipated policy benefits of the initiatives against the costs posed by cartels on consumers, the competitive process and the economy. Where possible, the policy initiatives should include mechanisms that preserve and protect the competitive process.

Chapter 3

The pricing dynamics of a multi-product cartel

3.1   Introduction

Having identified, in the previous chapter, that a significant number of cartels detected and prosecuted in South Africa involved multi-product collusion, in this chapter panel data is used to explore the pricing dynamics of a multi-product cartel with known phases of instability, though the exact duration and timing of those phases may be less certain. The wire cartel is used as a case study with the aim of better understanding the estimation of overcharges on different products, in a context where prices for the various products were set within the same collusive framework. Cartel overcharges can vary by product for reasons that include differences in conditions of demand and supply. This raises questions for competition authorities and courts when judging overcharges by a multi-product cartel, which may be a consideration in determining penalties or in damages claims.

Typically, firms that participate in a multi-product cartel face a single investigation for all the products that they cartelise and will each pay a single penalty relating to all the cartelised products. For instance, concerning the wire cartel, a single case was formulated for all its cartel-related conduct, arguably because this was perpetrated by the same firms, despite it covering multiple products and taking different forms. In many instances, firms that participate in multiple cartels pay a single penalty for multiple cases of cartel conduct. The question that arises is whether the competition authorities and courts should rely on an average of overcharges on different products or treat each product on its own merits and find a pragmatic rule for determining the overall penalty. Addressing this question requires developing a deeper understanding of the overcharges at the product level, since the time-series properties of aggregate data (across all products) may differ significantly from those at the product level.

The wire cartel offered some useful insights into variations in cartel overcharges over the life of the conspiracy, because of changes in demand and supply conditions and in the dynamics of the cartel itself over its lifespan. The cartel period represents the typical lifespan of a cartel, proceeding through phases of stability, instability, readjustment following a period of instability, and eventually breakdown. The cement cartel, which involved multiple products, went through similar phases as it adjusted after the end of the legal cartel in 1996 [see Boshoff and Van Jaarsveld (2019)]. A price war followed the end of the legal cartel. The price war was

succeeded by a period of re-establishment and, subsequently, a sustained period of collusion. Similarly, the wire cartel was characterised by intermittent periods of instability. As argued below, this raises questions about the appropriate technique to use in determining overcharges from such a cartel. For instance, the use of a before-and-after approach, which fails to account for the different phases, may produce biased overcharge estimates.

The ability of cartels to raise prices above the competitive level varies over the term of the conspiracy: collusive prices may fluctuate depending on cartel efficacy. Estimating overcharges at different times allowed the author to explore temporal variations in cartel overcharges. It also presented an opportunity to account for the presence of phases of instability during the life of the conspiracy. The wire cartel provided fertile ground for developing practical insights and understanding some of the issues in overcharge estimations. It also helped highlight the drawbacks of approaches that assume that a cartel's pricing power remains the same throughout its lifespan. The presence of unstable periods of collusion has practical implications for the courts handling damages claims and competition authorities wishing to use overcharges to demonstrate their value-add.

The author used panel data and the during-and-after[45] technique, with dummy variables capturing the cartel effects. As shown by Boswijk, Bun and Schinkel (2019), misdating the beginning and ending of a cartel has serious consequences for estimated overcharges, and this will also apply to misdating the beginning and ending of the periods of instability. The stable collusive periods were captured using dummy variables for the different cartel phases, taking the value of 1 during the phases of collusion and 0 otherwise. The author controlled for known periods of instability (which cartel members commonly termed the price war period), using a separate dummy variable for those episodes that allowed for the isolation and capture of the *pure* cartel effect. In the study three alternative assumptions about the transition period between collusive and non-collusive episodes were made:

(i)   First, it assumed that a transition period between collusion and competition could be captured using a separate dummy variable.

---

[45] This is the before-and-after technique, but the study took the period after the conspiracy as being competitive. The study acknowledged that there is significant debate about the appropriateness of using the post-period as a competitive benchmark, due to the challenges associated with identifying the ending of the cartel and its effects. This period was taken because of the availability of data, and different dates were taken as the ending of the conspiracy, which allowed us to undertake a sensitivity analysis.

(ii) The second assumption made in the study was that there was a linear transition period between collusive pricing and competitive pricing. A separate indicator variable was used for the unstable periods, and different cartel phases were captured using separate dummy variables for each phase.

(iii) A third assumption made was that there was no transition period, so that the cartel instantly adjusted from pricing at collusive levels to pricing at competitive levels. Again, a separate indicator variable captured the unstable phases and separate dummy variables were used to capture the individual phases of collusion.

Panel-data analyses enabled the study to capture the dynamics of collusion across the multiple products involved. Each dummy variable that captured the cartel effect in each phase allowed for observation and comparison of the magnitude of overcharges in each phase relative to the other phases. This enabled the author to observe temporal variations in overcharges. As discussed below, the results for the wire cartel show that overcharges varied significantly by products and over time, with some products having demonstrably low and statistically insignificant overcharges. In other products, the measured overcharge decreased significantly from one collusive phase to the next. Overall, the results reveal that the cartel under study was imperfect, and cartel members potentially faced significant competition for some products within the collusive framework (e.g. from a competitive fringe). The results also show that the cartel optimised its pricing on different products in line with the constraints it faced.

This chapter, therefore, contributes to the economic literature on overcharge estimation for (i) multi-product collusion, (ii) multi-period collusion where the price effects vary across collusive phases; and (iii) where transitions between collusion and competition can take different forms. The next section reviews the literature related to these three themes.

3.2   Review of literature on multi-product pricing, collusion, and overcharges

This section presents a discussion of three dimensions of the economic literature on cartel pricing. First, the author selectively reviewed the economic literature on the pricing behaviour of multi-product firms, focusing on multi-product firms involved in collusion. When firms produce multiple products, they can be expected to implement pricing in a way that optimises cost recovery, revenue and profitability across their entire product portfolio, and the same considerations may remain relevant in a cartel setting. As noted earlier, this literature is relevant to the study because the wire cartel entailed the production and cartelisation of multiple

products. Second, the author surveyed the economic literature on methodologies for measuring overcharges, focusing on those that were relevant to the approach adopted in this chapter. Third, given that the post-cartel period was adopted as the counterfactual period, the author also reviewed the literature on how to account for the transition from collusion to competition following the collapse of a cartel. In this regard, the importance of correctly dating the end of collusion was considered.

The review was focused on the international literature, but it included references to overcharge studies based on South African cases. The approaches applied in the South African studies fell within the same classes of techniques as those applied in international empirical work.

3.2.1 Brief review of literature on the pricing behaviour of multi-product firms

Firms in most markets produce and supply multiple related products that share elements of the production process, e.g. production equipment, production space, staff members, among others. Further, where a firm produces multiple products, some products are likely to be value-added product extensions of base products. In such markets, the optimal pricing strategy may be to ensure that overall production costs (including common costs) are recouped throughout the entire portfolio of products [O'Donoghue and Padilla (2013)]. This necessarily means that the firm earns different margins for products facing demand with different price elasticities.

O'Donoghue and Padilla (2013) suggest that in such circumstances, it might be more insightful to analyse the pricing policy of a multi-product firm in its entirety and not in a piecemeal fashion, product by product. Although argued in the context of excessive pricing, it appears that the same reasoning might be relevant to a setting in which a firm involved in collusion is faced with given costs and produces multiple products that face different demand and supply conditions. The pricing decisions of the firm are potentially informed by the need to optimise the recovery of costs or maximise revenue generation, which in turn enables the firm to optimise and generate the highest possible aggregate profit. In that setting, the ability to impose overcharges can be expected to vary by product. The presence of a competitive fringe, which affects some, but not all products, can also affect the magnitude of cartel overcharges on selected products.

Harrington (1987) examined the feasibility of collusion involving multi-product firms where the products were related. The study examined whether the output decision of a firm in one product or market was affected by the past behaviour of rivals in another product or market,

over and above the past behaviour of its rivals of the product, or in the market in question. Harrington argued that this is plausible where two products share cost or have demand dependencies. The study also showed that collusion is possible because of the existence of credible punishment extending to another product or market, as well as the possibility that collusion in the product of concern may also cease, leading to reversion to competition and thereby raising the cost of defection.

Bernheim and Whinston (1990) studied the relationship between multi-market contact and collusion in a repeated game setting, drawing three key conclusions. First, multi-market contact can enhance collusion. Second, firms benefit from multi-market contact. Third, the impact of multi-market contact is not necessarily socially undesirable.More significant for this chapter, the authors found that the impact of multi-market contact on prices or profits in any one of the contact markets was significantly influenced by the nature of the markets and the features of the firms involved, as well as by potential entrants. The authors argued that when firms face cost asymmetries, prices may be higher or lower depending on discount factors, whereas with similar firms, prices and profits may be higher in some markets than in others. Multi-market contact leads to interdependent behaviour across markets, meaning that when an event occurs in one market, correlations can be observed with prices in other markets.

Extending the work of Bernheim and Whinston, Spagnolo (1999) argued that there exists another circumstance in which multi-market contact always facilitates collusion. This occurs when strategic interactions between firms are interdependent, with the consequence that a firm's valuation of profits in one market is influenced by profits achieved in other markets. Punishment is more effective, and collusion is more profitable when multi-market contact exists. Spagnolo concurred with the findings of Bernheim and Whinston's study, arguing that when the circumstances identified by them exist for collusion to arise from multi-market contact, the pro-collusive effect is stronger. Matsushima (2001) showed, at a theoretical level, that multi-market contact facilitates implicit collusion, even under imperfect monitoring. This is especially the case when firms' supply choices in individual markets depend on developments in those markets and developments in other markets.

Using collusion in airlines, Ciliberto and Williams (2014) empirically studied the relationship between multi-market contact and collusion. The authors found that airlines that interacted in multiple markets co-operated in setting fares, which was consistent with the findings from the theoretical literature, which shows that multi-market contact facilitates collusion. On the other

58

hand, the authors found that airlines that did not enjoy multi-market contact did not co-operate on fare-setting behaviour. The impact of multi-market contact on prices was significantly affected by cross-price elasticities.

As such, the presence of multi-product or -market contact, especially where products are related, makes collusion more likely to involve a wider range of products or markets. This factor has been widely flagged in the economic literature as possibly facilitating collusion. Chapter 2 provides a discussion of the significance of this factor in South African cartels and shows that it has been a feature of collusion. Cartels such as those in the wheat-to-bread and steel value chains involved the same firms colluding on several products and markets. This is supported by the evidence presented in Chapter 2. The World Bank (2016) found that 48 of the collusion cases that had been uncovered involved a firm that had colluded in another market in the same sector or in a different one, and for at least 20 of the cartels where there was a leniency application, the leniency application came from a firm that had applied for immunity for multiple cartels.

The theoretical literature cited above shows that multi-product cartels recognise inter-dependencies between products and markets and take that into account when making pricing decisions. However, there is limited empirical work examining the way multi-product firms involved in collusion approach the pricing of individual cartelised products. This study contributes to the empirical literature in this aspect.

3.2.2 Brief review of literature on multi-period collusion

Apart from the many cartels operating across multiple products, evidence from an examination of the duration of cartels shows that most cartels operated over multiple periods [see, for example, Chapter 2, above; World Bank (2016); and Levenstein and Suslow (2006; 2011)]. Macroeconomic conditions, conditions of demand and supply, and firm-specific financial circumstances fluctuate over the duration of an existing cartel. As these fluctuations occur, firm incentives to honour the collusive agreement may shift, leading to deviations from the collusive equilibrium. These deviations could lead to a shift to a non-collusive equilibrium. Cartels that survive the impact of these fluctuations adapt their conduct to the circumstances, in some instances investing in mechanisms to keep incentives intact and making compromises where necessary (e.g. accepting a lower overcharge in one period compared with another). Adaptation can be expected to affect the extent to which cartels are able to exercise market power

throughout their lifespan. Briefly discussed below is some literature on the relationship between fluctuations in macroeconomic conditions, demand and supply conditions, and the financial condition of firms and cartel success.

Green and Porter (1984) predict that a negative shock to demand that is not observed will result in a price war, while Rotemberg and Saloner (1986) show that price wars will arise during booms. This suggests that changes in economic and market conditions result in changes in the pricing behaviour of firms, including when they are involved in a cartel. Markets experiencing moderate growth are more likely to support collusion than those that are either rapidly shrinking or rapidly growing [Symeonidis (2003)]. Rapid growth increases the chances of entry and raises the level of uncertainty. Jacquemin, Nambu and Dewez (1981) made similar findings on the impact of higher demand growth on the duration of a cartel. Any given market can be expected to go through various phases, growing moderately or rapidly at some point, or experiencing decline as the macroeconomic and demand environment may dictate. Eckbo (1976) examined the relationship between demand elasticity and the success of cartels and found that cartels' success in raising prices significantly was possible if demand was inelastic and the cartel faced limited threats from substitutes in the short term.

The ability to raise prices can vary as substitutes or new entrants emerge over time. The threat of entry is a significant contributor to cartel death [Levenstein and Suslow (2006;2011)]. The ability of a cartel to institute entry deterrence successfully over time will also affect its ability to raise prices significantly over the course of its lifespan. If the threat of entry increases in time, the ability of the cartel to significantly raise prices over multiple periods can be expected to progressively decline.

Similarly, Levenstein and Suslow (2011) argue that fluctuations in the discount rates of firms affect the prospects of cartels' success. Firms facing financial challenges may become impatient, causing them to deviate from the collusive equilibrium. As such, fluctuations in discount rates affect the pricing behaviour of firms during the life of the cartel. A firm facing financial distress may also not be as concerned about punishment for deviation, given its more immediate need for cash compared with waiting for collusive rents, which may not be sufficient to keep it afloat.

Although this section places the emphasis on some literature relating to macroeconomic conditions, demand and supply factors, the impact of fluctuations in firm-specific discount

rates on the incentives for firms, and the threat of entry, other challenges facing the cartel could also affect the ability of firms to continually exercise cartel market power. These include imperfect information, which may enable episodic cheating to go undetected. Furthermore, if the threat of cartel detection by the competition authorities increases over time, a cartel may progressively adapt its pricing to avoid detection, meaning that it may realise varying overcharges over time. Harrington (2004) discussed cartel pricing dynamics when there is a risk of detection by a competition authority and finds that cartel prices may be lower if sanctions are endogenous and the probability of detection is fixed. In that case, the cartel price path decreases over time, after being initially raised. The implication of this is that one can expect variations in overcharges over time.

The theoretical literature predicts varying and even episodic overcharges, but there are limited empirical studies, especially involving multi-product cartels. The few studies that exist include Boshoff and Van Jaarsveld (2019), and Boswijk, Bun and Schinkel (2019).

3.2.3 Synopsis of classes of approaches to overcharge estimation

Firms collude to raise prices above the level that would exist absent the conspiracy, which is called the *but-for* or counterfactual price level [Motta (2004)]. Cartel overcharges attempt to estimate the extent to which prices charged by the cartel were higher than the *but-for* price. The biggest challenge in estimating overcharges is determining the *but-for* price, which is not observed. The economic literature identifies three broad classes of techniques that researchers can apply to estimate the *but-for* price and consequently the cartel overcharge. These are comparator-based approaches, financial-analysis-based approaches, and market-structure-based approaches [Niels, Jenkins and Kavanagh (2011) and Hüschelrath, Müller and Veith (2013)].

Comparator-based approaches are sub-divided into cross-sectional, time-series (or temporal), and difference-in-differences (DiD), which combines cross-sectional and time-series elements. The counterfactual in comparator-based approaches is based on comparable, but non-cartelised markets or a competitive period in the same market, or both in the case of DiD. Financial-analysis-based approaches apply financial information such as profitability and costs for firms involved in the conspiracy and those of comparable firms and markets not exposed to collusion to estimate the counterfactual [Niels, Jenkins and Kavanagh (2011) and Rubinfeld (1985)]. The application of industrial organisation models of competition, in combination with empirical

61

estimation to estimate the counterfactual price fall into the last class of approaches. Profitability is difficult and complex to measure and, as such, there is limited application of financial-analysis- and market-structure-based approaches [Hüschelrath, Müller and Veith (2013)]. This study was primarily concerned with techniques falling within the class of comparator-based approaches, briefly discussed below.

3.2.4 Before-and-after approach and estimating overcharges

Time-series techniques use information relating to the cartelised product(s), covering the period of conspiracy and a period when the cartel did not exist, to determine the level of the overcharge. The period without the conspiracy can be a period prior to the conspiracy or the period following the end of the cartel. This is often referred to as the before-and-after technique. In its simplest form, the before-and-after technique involves simple comparisons of prices between the conspiracy and non-conspiracy periods. The challenge with this approach is that the entire difference between the cartel price and the *but-for* price is attributed to the overcharge, when it could be partially a result of demand and supply shocks [White, Marshall and Kennedy (2006)]. To address this challenge, a more sophisticated multivariate approach is preferable [Connor (2008) and Paha (2011)].

There are two options for estimating the cartel effect, which should ideally be used as complements to ensure the robustness of analysis [Nieberding (2006)]. First, a regression model can be estimated for the non-conspiracy period, and the resulting coefficients are used to estimate the *but-for* price during the cartel period (often referred to as the forecasting or backcasting approach). For robust results, this approach requires sufficient observations in the non-conspiracy period to ensure the reliability of the parameters. Alternatively, the dummy variable approach can be applied, taking the value of 1 in the conspiracy period and 0 otherwise, with the coefficient of the dummy variable measuring the cartel effect. The use of one scale dummy[46] may not appropriately capture the cartel effect when the cartel affects prices in varying and complex ways [Finkelstein and Levenbach (1983)].

In both the forecasting and dummy variable approaches, the start and end dates of the conspiracy must be captured as accurately as possible, as misdating has a significant impact on the accuracy of overcharge estimates [see, for example, Boshoff and Van Jaarsveld (2019); and Boswijk, Bun and Schinkel (2019)]. This requires a close examination of the facts of the case

---

[46] Assumes that the cartel added a fixed amount or percentage to a price.

at hand and making the necessary adjustments to one's assumptions [ABA (2010)]. When the dummy variable is defined at the same frequency as the data in the estimation sample, the dummy variable and the forecasting approaches are equivalent [Higgins and Johnson (2003)]. This equivalence occurs when a separate dummy variable is included for every month/quarter/year during the cartel period.

Before-and-after techniques are easier to apply than other techniques and have limited data requirements. They mostly require time-series data on the cartelised product, under the assumption that the behaviour of the firms in the collusive period would have been similar to their behaviour in the non-collusive period had the cartel conduct not taken place. Identifying the starting and ending of cartel effects on price is a key challenge with the before-and-after approach. For instance, where the cartel period is followed by tacit collusion, prices will not reflect competition, leading to a downward bias in estimates of cartel overcharges [Harrington (2004), Connor (2008) and Boswijk, Bun and Schinkel (2019)]. This is also the case where the market involves long-term contracts that are not necessarily terminated or renewed at the time the cartel is uncovered or prosecuted.

The difficulty of accurately identifying the end of collusion and its effects on prices can be achieved by allowing for a transition period between collusion and when competition can be expected to be reflected in prices [Hüschelrath, Müller and Veith (2013)]. If the end of collusion is characterised by a price war, cartel overcharge estimates might be biased upwards [de Coninck (2010) and Frank and Schliffke (2013)]. To avoid these challenges, a counterfactual period prior to the conspiracy is preferable [Van Dijk and Verboven (2008)]. In South Africa, this counterfactual is not available in some markets because of the history of co-operation and legal cartels [see for example, Boshoff (2015)]. This applies to the wire cartel, which had a history of co-operation prior to the 2001 agreement.

Despite the substantial number of cartels uncovered in South Africa, the body of empirical research and the application of quantitative techniques to estimate cartel overcharges is still limited. The studies in South Africa include those using spatial techniques to estimate overcharges by the bitumen cartel, difference-in-difference techniques applied in the precast concrete products cartel and temporal techniques in the wheat flour cartel.

Boshoff (2015) used spatial econometric techniques to estimate overcharges in the bitumen market, which was previously characterised by a legal cartel that was succeeded by an illegal

63

cartel. In this case, before-and-after techniques were not appropriate. Thus, the author relied on a panel analysis, using comparator markets to estimate parameters for the long-run relationship between price and explanatory variables. He used these parameters to forecast the *but-for* price for South Africa, and then compared this price to the actual price charged by the cartel, to derive the overcharge. He also applied the dummy variable technique, by including the cartelised market in the panel and including a dummy variable in the specification, to capture the cartel effect. The dummy variable approach has the advantage of demonstrating how explanatory variables differ in their impact on the cartelised market compared with other countries and estimates the difference in price levels across countries. The author also applied a DiD analysis, in which he included a cartel period dummy, a dummy for the cartelised market (South Africa) and an interaction term, which estimated the overcharge. The DiD analysis yielded a much higher overcharge (80 per cent) estimate than did the forecasting and dummy variable approaches employed (25 per cent).

DiD analysis was also applied to estimate overcharges in the precast concrete products cartel [Khumalo, Mashiane and Roberts (2014)]. In this approach, the authors identified comparator products, for which they calculated the ratio of prices between the cartelised product and the comparator products during a competitive period. They used these ratios to draw inferences about the counterfactual price and consequently the cartel overcharge. This is a simple approach which assumed that the cartelised product and the comparator products faced the same shocks, such that the ratio was stable. More sophisticated techniques may be necessary in circumstances where this assumption is not reasonable. Khumalo, Mashiane and Roberts (2014) found that overcharges in the concrete pipes cartel ranged between 16.5 per cent and 57 per cent, depending on the region.

The wheat flour cartel was uncovered in 2007. To estimate overcharges by this cartel, Mncube (2014) used the dummy variable approach in a multivariate regression model, in which the period following the end of the cartel was identified as the competitive period. The market was regulated prior to the illegal cartel. Mncube (2014) found overcharges ranging between seven per cent and 42 per cent. The potential limitations of the before-and-after approach led Mncube (2014) to supplement the results with a review of industry profits over time.

Using the cement cartel, Boshoff and Van Jaarsveld (2019) observed that cartels experience different phases when periods of collusion may be intermingled with periods of competition. To examine the impact of this on overcharge estimates, the authors used a Markov regime-

switching model and found that the model reported higher overcharge estimates than those reported by conventional models. Estimated overcharges ranged between 19.2 per cent and 19.9 per cent. The authors argued that a regime-switching model enables better estimates of cartel overcharges because it (a) accounts for differences in the data-generating process during collusive and non-collusive episodes, (b) permits the dating of such episodes, and (c) allows for the possibility of flexible transitions between collusive and non-collusive episodes.

As noted previously, empirical studies of interrelated multi-product overcharges are limited, and there is limited focus, from a portfolio perspective, on the pricing behaviour of multi-product cartels. Similarly, as noted previously, there is limited empirical literature on the temporal variation of cartel overcharges when collusion took place over multiple periods. This chapter presents an examination of the temporal variation of overcharges by a multi-product cartel that existed over multiple periods, which were interspersed by periods of instability. One aspect of the empirical treatment involved assumptions about the nature of the transition between collusion and competition. This is a further dimension in which this chapter improves on the existing literature on estimating cartel overcharges. To this end, the following subsection briefly reviews how the work by Hüschelrath, Müller and Veith (2013) has dealt with transitions between competition and collusion.

3.2.5 Cartel overcharges under different assumptions about the transition from collusion to competition

Hüschelrath, Müller and Veith (2013) applied a multivariate before-and-after technique to estimate cartel overcharges from the German cement cartel. In doing so, they recognised that even after the breakdown of a cartel, there may not be an instantaneous switch back to competition. This may be due to new tacit forms of collusion, medium-to-long-term contracts, and general price rigidities, all of which may result in the persistence of collusive prices. The authors suggested that econometric approaches must therefore explicitly account for the transition from collusion to competition, but without overstating it. Assumptions about the form and nature of the transition period will impact on the estimated cartel overcharges. As such, it is important to understand how long, and up to which point in time the collusive price effects persisted, as well as the functional form of the transition period [Hüschelrath, Müller and Veith (2013)].

Three options to account for the transition period are identified by Hüschelrath, Müller and Veith (2013). The first entails ascribing the full cartel price and consequently overcharge to the cartel and transition periods, which is achieved by taking the cartel dummy variable as 1 during both periods and 0 otherwise. The second assumes that cartel effects gradually dissipate, meaning the cartel dummy variable linearly declines from 1 to 0 during the transition period. The third includes an additional indicator variable that accounts for the transition period, by having a value of 1 in the transition period and 0 otherwise. Hüschelrath, Müller and Veith (2013) found higher overcharges after allowing and controlling for a transition period than they found when assuming an instantaneous adjustment from collusion to competition.

This study followed their example, by looking at transition assumptions in the context of a multi-product cartel. This is important because, even in a multi-product setting, assumptions about the transition between collusion and competition affects overcharge estimates. Some cartels degenerate into a price war characterised by sub-competitive pricing. If the overcharge estimation technique is not adjusted to control for the price war phase, the resulting overcharges will possibly be overestimated. In those cartels where there may be persistent cartel price effects that dissipate gradually before prices reach competitive levels, overcharges will be under-estimated.

3.3   Background to the wire industry and the cartel

Steel-based products are often intermediate products that are used in several downstream markets spanning sectors. Some of the products can be further processed to produce other value-added products. Steel is the primary input used in the production of the cartelised products. Other inputs include zinc and labour. The products are produced by several related producers and some independent producers, but they can also be imported. The four largest firms produced and supplied a broad range of these products across the country while others supplied a limited range of products, in some instances, in specific geographic areas. The members of the cartel operated across the major product lines. Steel products are homogenous, with firms competing primarily on price.

Toward the end of 2009, the Commission referred a case to the Tribunal, alleging that the major suppliers of some wire products had colluded by fixing prices among other collusive practices. The case ran for a protracted period, spanning over 10 years because of procedural issues. The larger firms resolved cases against them with the Commission. One of the leading firms

obtained leniency; the other settled with the Commission in 2018; and another, smaller one, settled after that. The other members of the cartel are still being prosecuted. As shown in Chapter 2, collusion was systemic across the entire industry, and it was characterised by mutually re-enforcing arrangements such as joint ventures, cross-shareholdings, cross-directorships, and *quid pro quo* arrangements. This study focused on the price-fixing element of the cartel, which took the form of list price agreements, supported by agreements on levels of discounts for different classes of customers.

The cartel monitored the behaviour of members, using various mechanisms, which included self-reporting mechanisms that exploited customers' desire to bargain. This included contract clauses that required customers to provide evidence of written offers from rivals to their supplier and afford the supplier an opportunity to match the offers. In other instances, customers who received better offers informed their suppliers of these offers, requesting them to match or better the offers, or risk losing business. Despite being elementary, these forms of monitoring effectively detected cheating, as firms would use them to address cheating. A firm would thus know if it had lost one of its traditional customers, and in some instances, to whom that customer had been lost.

Many cartels experience cheating during their lifespan, and this cartel was no exception. As occurs in many cartels, cheating may result in periods of temporary movement to pricing below cartel levels. This period can be in the form of an intense price war or tit-for-tat pricing – in which the response to cheating is simply to match offers made by rivals (which does not necessarily reflect an intense sub-competitive price war) – or some other form of pricing that is in the intermediate range between deep price wars and cartel prices, but does not necessarily reflect competitive pricing, as prices could still be above competitive levels.

The conduct lasted for a sustained period spanning potentially more than a decade, predating the current competition law regime. The Commission considered the cartel conduct to have started in the period from 2001.

3.4 Data and methodology

3.4.1 Data

The study used data on the average pricing of the cartelised products, together with data on the key cost drivers associated with manufacturing these products (steel and zinc) – which affected

the final prices – and an indicator of demand for the respective products. Data on steel and zinc prices was obtained from international sources, particularly Index Mundi. Data on the economic performance of the steel sector, sectors that contributed to the demand for the cartelised products, and appropriate price indices were obtained from Statistics South Africa (StatsSA) and Quantec. In general, South African steel prices and consequently those related to steel-based downstream products followed the international price of steel (see Figure 3.1). This was largely because the major domestic producers of primary steel, such as Arcelor Mittal, had historically applied an import-parity pricing strategy (see, for instance, Tribunal case number 13/CR/Feb04). Steel is a major input in the production of the cartelised products, accounting for a significant proportion of production costs.

Figure 3.1: Domestic PPI for billets and mild steel, and international steel prices over time
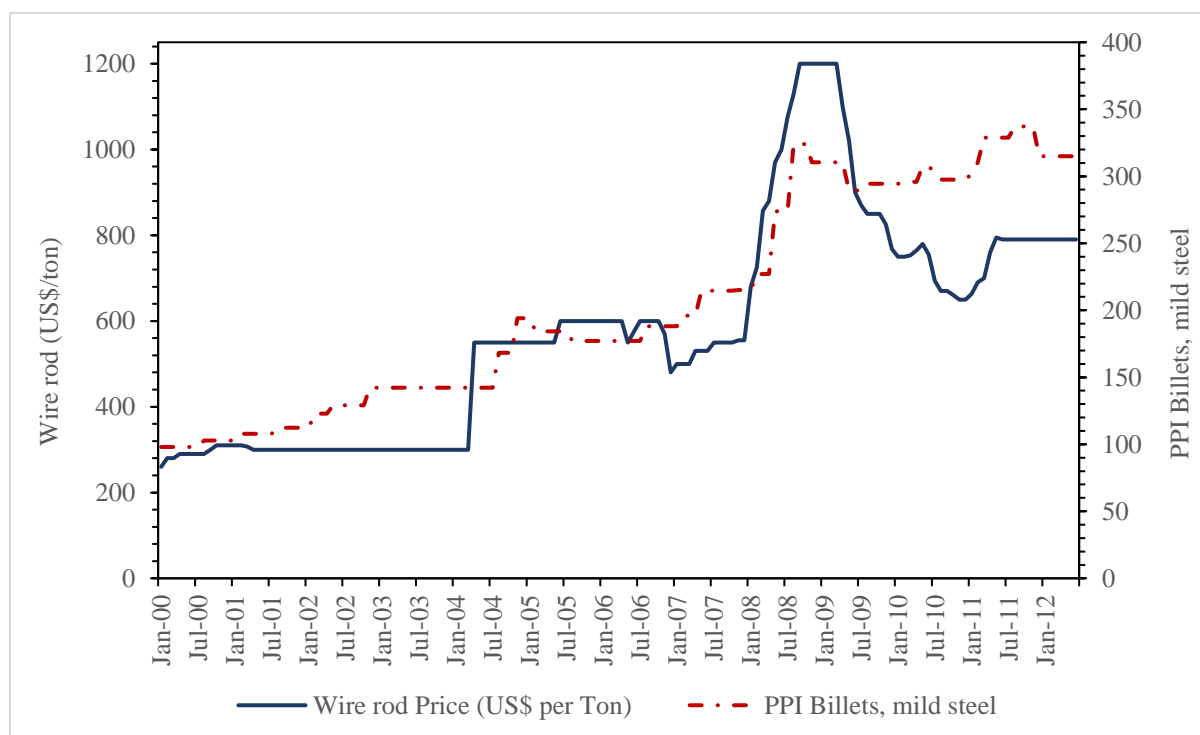
Source: Index Mundi and StatsSA PPI data

The wire cartel operated for a sustained period, which was suspected to have started prior to the introduction of the current competition law regime. The Commission's investigation was focused on the period after the introduction of the current legal regime. Witness evidence from the case suggests that there were three collusive phases, as follows: the first phase took place prior to the 2001, ending with a period of competition; the second phase took place between 2001 – when the cartel re-established itself – and 2005, when a price war erupted (the price

war occurred between 2005 and 2006); and the third phase took place between 2006 and 2008, when the cartel was ended by the Commission's intervention. For the purposes of examining overcharges from multi-period collusion, the study focused on the second and third phases. Figure 3.2 shows the cartel gross margin calculated as price over the cost of key direct materials (wire rod and zinc). Figure 3.2 shows that during the price war period there is an observable decline in the gross margins across several products.

Figure 3.2: Gross margins by product (price less cost of key direct materials)
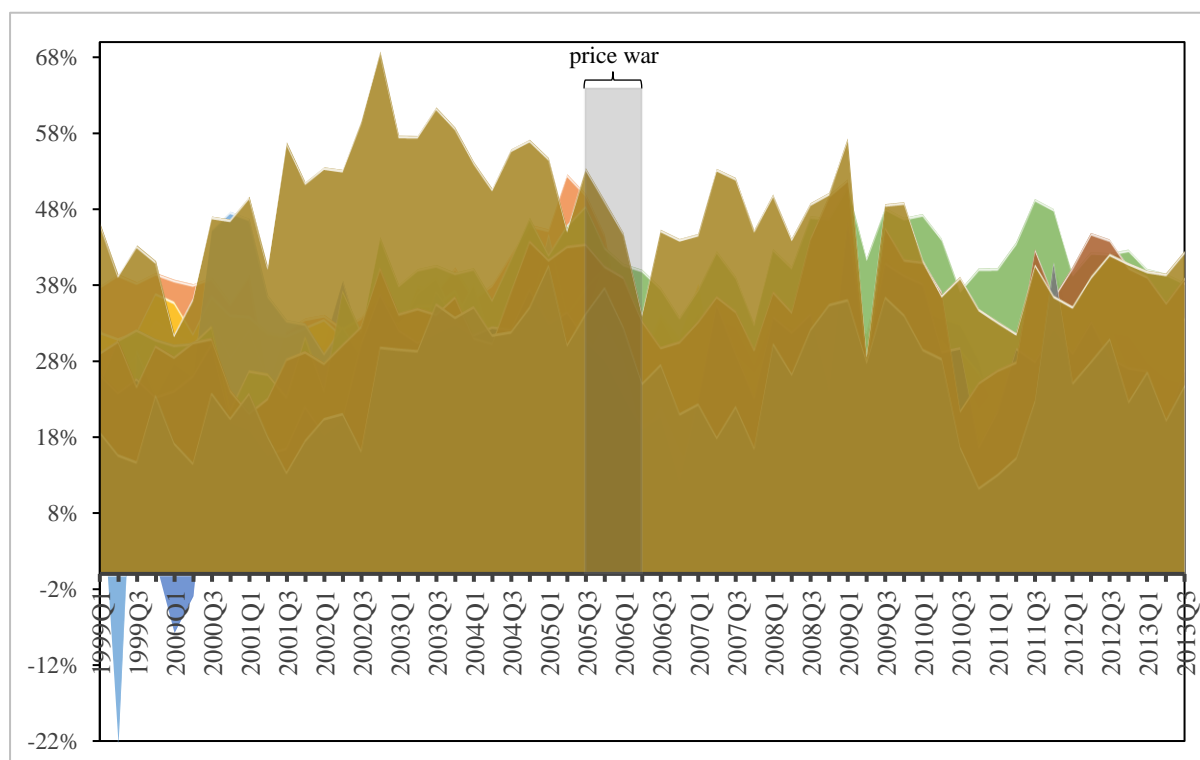
Source: Own calculations based on product data

Taking these events into account, Table 3.1 reports the variables used in the analysis.

Table 3.1: Variables

Variable	Description	Source
Price (P)	Average price per tonne of the cartelised products	StatsSA
Steel	Price per tonne of wire rod for individual steel-based products	Index Mundi
Zinc	Price per tonne of zinc for individual steel-based products	Index Mundi
Gross Domestic Product (GDP)	Gross domestic product for South Africa	Statistics South Africa/Quantec
Cartel phase dummy	Indicator variable to capture individual cartel phases for each product	Constructed to take the value of 1 in cartel phase and 0 otherwise

69

Variable	Description	Source
Price war dummy	Indicator variable to capture price war effects for each price war phase per product	Constructed to take the value of 1 during price war phase and 0 otherwise
Transition period dummy	Indicator variable to capture transition period at the end of the cartel, per product	Constructed to take the value of 1 in transition period and 0 otherwise

Source: Own descriptions based on regression data set

For the purposes of the analysis, the quarterly data formed a balanced panel of 10 cartelised products over 14 years. Although import competition could potentially have undermined the cartel, it must be noted that the cartel members produced and supplied the bulk of the domestic consumption of the products. There was also no evidence of imports having formed a substantive basis of pricing considerations among the firms during the cartel period. Table 3.2 provides key summary statistics of the data used in the analysis, which includes the per-unit price of products, per-unit input costs for steel and zinc, and GDP (in millions of rands).

Table 3.2: Summary statistics

Variable		Mean	Std. Dev.	Min	Max	Observations
Price	Overall	6,738.00	3,439.96	1,917.20	19,324.77	N = 590
	Between		1,673.08	4,250.77	9,081.61	n = 10
	Within		3,051.19	1,400.05	17,450.71	T = 59
Steel	Overall	4,303.61	2,077.58	1,296.00	10,751.00	N = 590
	Between		823.28	3,034.00	5,435.37	n = 10
	Within		1,924.91	1,386.24	9,619.24	T = 59
Zinc	Overall	178.83	326.10	0.00	1,854.86	N = 590
	Between		269.98	0.00	743.33	n = 10
	Within		201.57	-232.50	1,290.36	T = 59
GDP	Overall	467,319.80	200,906.70	194,211.20	844,284.30	N = 590
	Between		0.00	467,319.80	467,319.80	n = 10
	Within		200,906.70	194,211.20	844,284.30	T = 59

Source: Own calculations based on regression data

Figure 3.3 shows the average prices for each product and the average weighted price across all products. There was significant variation between the average weighted price and the products' prices, as well as across the prices of the individual products. The average weighted price conceals these variations.

70

Figure 3.3: Index of average prices
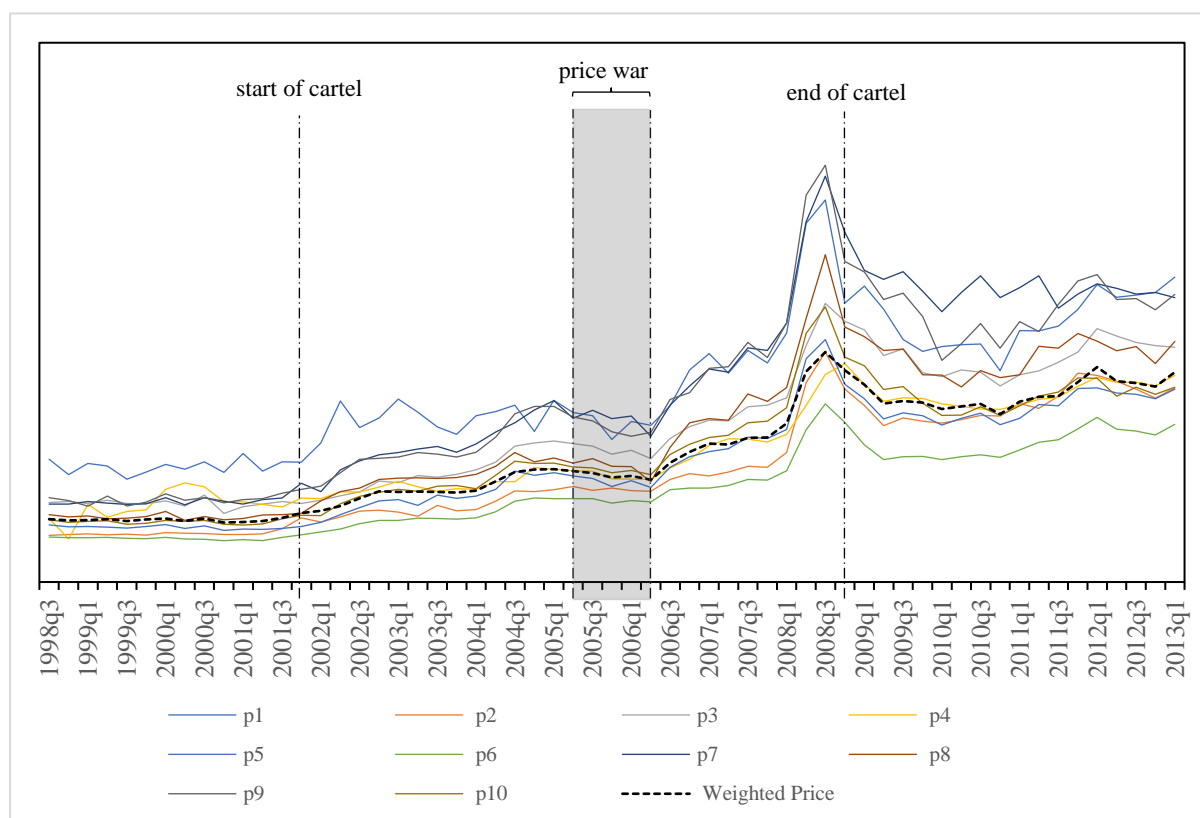
Source: Own calculations based on product data

3.4.2 Econometric approach

The study used the before-and-after technique to estimate cartel overcharges. The counterfactual period was taken as the period following the end of the cartel. This approach was appealing because it used data on the cartelised products, and it avoided the difficulties of identifying suitable comparators for the cartelised products in South Africa. There was some evidence suggesting that the cartel existed prior to 2001, when the price fixing framework was established. Between 1998 and 2008, the cartel experienced known periods of instability in the form of a non-collusive phase leading up to cartel re-establishment in the third quarter of 2001, and a price war that took place between the second quarter of 2005 and the second quarter of 2006. Figure 3.3 shows a downward trend in prices during the price war period. These two phases are significant because they necessitated significant meetings aimed at resolving the cartel's challenges. The two meetings (around the third quarter of 2001 and towards the second quarter of 2006) give a clear indication of when co-operation began in each phase. Figure 3.2 shows that following these meetings, prices increased, trending upwards during the two collusive phases that occurred between 2001 to 2005, and between 2006 to 2008 (final quarter of 2008), when the cartel ended. Debate about the exact dating of the end of the collusive

71

phases can arise because this might affect estimated overcharges. Evidence from the case, taken together with the price evidence presented in Figure 3.3 enable us to date the phases accurately.

The study tested the individual price series for structural breaks, using a fluctuation test for a significant shift in average price changes over time [Crede (2019) discussed various approaches]. This was augmented by the Bai and Perron (1998, 2003) technique for dating structural breaks, which was carried out by regressing each price series on a constant. It can identify points when there were significant changes in the data-generating process. Crede (2018) used this test to check the robustness of structural break tests, in order to identify and date collusive phases in European pasta markets. The procedure was similar also to the price-change approach used by Hüschelrath and Veith (2013) to screen and date collusion in German cement markets. The results of the structural break tests were complemented by the facts pertaining to key events concerning the operation of the cartel and changes in key production costs. Table 3.3 presents the test results.

Table 3.3: Results of structural break tests

Variable	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Ho: There is no structural change in the series										
F-stat	172.28 ***	295.61 ***	249.24 ***	292.58 ***	229.60 ***	212.14 ***	248.62 ***	263.05 ***	199.51 ***	204.70 ***
pvalue	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BP 1	2002q1	2001q3	2002q1	2002q1	2002q3	2002q2	2002q1	2002q1	2002q1	2002q1
BP 2	2006q4	2004q1	2004q1	2006q1	2006q1	2004q2	2004q2	2006q1	2004q2	2004q2
BP 3		2006q1	2006q1	2008q1	2008q1	2008q1	2007q4	2008q1	2007q4	2007q4
BP 4		2008q1	2008q1			2011q1	2009q4		2009q4	2009q4
BP 5		2011q1								

*Statistically significant at: 1 per cent (p<0.01) ***, 5 per cent (p<0.05) **, 10 per cent (p<0.10) **

The results show that the identified break points (BP) align with the cartel activity and shifts in costs. Breaks in late 2001/early 2002 and around 2006 correlate with the establishment of the two collusive phases identified above, whereas those breaks around late 2007/early 2008 and late 2009 and early 2011 correlate with shifts in steel costs. The gradual transition between collusion and competition could explain the lack of breaks associated with the end of collusive phases. Issues related to the transition between collusion and competition are discussed next.

The study considered two assumptions about the end of a cartel: first, instantaneous adjustment from collusion to competition; and second, a transition/adjustment period lasting four quarters following the initiation of the investigation. This is because prices changed quarterly, and there

were at least suggestions that communication between the firms continued in the quarter following the initiation of the investigation by the Commission. Further, allowing for a full year of absence of co-ordination allowed for cartel effects to dissipate more fully. Given these events, the author argues that there is a need to allow for an adjustment phase and to control the effects of these periods on overcharge estimates. Under the assumption of an adjustment/transition period, the study adopted two approaches. First, by using a separate dummy variable taking the value of 1 during the transition period and 0 otherwise; and second, by using a linear transition period from collusion to competition. This is in line with suggestions made by Hüschelrath, Müller and Veith (2013) on dealing with transition periods, as described earlier.

The dynamics of multi-product collusion and the resulting overcharges were captured using panel-data analyses, which has several advantages over either cross-section or time-series analysis [Hsiao, Chan and Ho-Illman (1995)]. First, it facilitates more accurate inferences about model parameters, by providing more degrees of freedom and more sample variability, thereby improving model efficiency. Second, it provides greater capacity to capture the complexities and thus enables better understanding of the dynamics of firm behaviour in multi-product collusion. It also enables one to control for the impact of omitted variables. Panel data therefore allows one to better understand dynamic relationships, including economic behaviour [Nerlove (2002)]. Using inter-individual differences reduces collinearity between current and lagged variables, which in turn, allows for the estimation of unrestricted time-adjustment patterns [Pakes and Griliches (1984)].

By pooling data, more accurate predictions of individual outcomes are possible [Hsiao, Chan, Mountain and Tsui (1989); and Hsiao, Appelbe and Dineen (1993)]. Panel analysis also provides the micro-foundations for analysing aggregate data. The time-series properties of aggregate data may differ from those of micro-level data [Granger (1990), Pesaran (2003)]. Where such differences are significant, aggregate estimates and resulting policy evaluations may be misleading [Hsiao, Shen and Fujiki (2005)]. This makes panel analysis suitable for understanding the dynamics of multi-product cartels, as opposed to generalising over different products by studying the dynamics of the cartel at an aggregate level. Finally, panel-data analyses can simplify computational and statistical inference in the analysis of non-stationary time-series data when cross-sectional observations are independent [Binder, Hsiao and Pesaran (2005), Im, Pesaran and Shin (2003), Levin, Lin and Chu (2002), Phillips and Moon (1999)].

Panel data allows for transformations that induce different and deducible changes in estimators, thereby resolving under-identification problems that would otherwise exist [Biøn (1992), Griliches and Hausman (1986), Wansbeek and Koning (1989)].

The relationship between price and its key drivers[47] can be explained by model 1, which can be estimated under three assumptions about the nature of the transition from collusion to competition. The three assumptions are (1) instantaneous adjustment, (2) a separate dummy variable capturing the transition phase, and (3) linear transition. In model 1, as follows, the current price does not depend in part on its previous values:

$$P_{it} = \alpha_i + \rho_i Demand_{it} + \beta_i Steel_{it} + \gamma_i Zinc_{it} + \sum_{j=1}^{j=3} \delta_{ij} D_{ijt} + \sigma_i trD_{it} + \omega_i pwD_{it} + \varepsilon_{it}, \qquad (1)$$

where P_{it} is the log price at time t for product i, and $Demand_{it}$ is the log for an indicator of demand, taken as GDP in this case, to avoid endogeneity problems associated with simultaneous determination of price and demand. Endogeneity can result in parameter estimates that are inconsistent, and potentially carrying the wrong sign. Using GDP is appropriate because some of the products were sold across multiple sectors. Further, given that the model that is finally applied (model 2 below) is in first differences with lagged dependent and explanatory variables, challenges of endogeneity between price and demand are mitigated. *Steel*$_{it}$ is log steel cost; *Zinc*$_{it}$ is log zinc cost; D_{jt} is an indicator of the cartel during different cartel phases *j=1* (prior to 2001, represented by D_1), *2* (between 2001 and 2005) and *3* (2006 to 2008); *trD*$_t$ is an indicator variable for the transition, and *pwD*$_{it}$ is an indicator dummy variable capturing the unstable (price war) periods; and ε_{it} are assumed serially uncorrelated error terms. The study was primarily concerned with cartel phases 2 (2001 to 2005) and 3 (2006 to 2008) and used a dummy variable D_1 to control for the period prior to the 2001 instability.[48] Under the assumption of instantaneous adjustment from collusion to competition, the transition dummy variable falls away. The cartel effect was measured by ($e^{factor} - 1$), where the *factor* equals the coefficient of the dummy variable.

Model 1, above, does not account for more complex pricing dynamics where the current price depends on the previous price(s). Nieberding (2006) suggests capturing this pricing dynamic by including a lagged dependent variable as part of the explanatory variables. However, the

---

[47] Evidence from the case suggests that prices were mostly driven by changes in steel prices, and this coincided with a lot of the cartel communication and price changes.
[48] There was not sufficient detailed information about the beginning of the instability during this period, so the study controlled for this period using a dummy variable.

current and past prices are still significantly correlated. Non-stationarity in time-series data poses the risk of spurious findings, especially for finite samples. Panel unit-root tests indicated that at least some of the panels contained unit roots, that is, they were non-stationary (see Appendix A). This suggests that there was a risk of observing spurious results. The author therefore tested for cointegration, using the Westerlund (2008) panel tests. There was evidence of cointegration (see Appendix A), meaning that the OLS estimates were highly consistent. The presence of cointegration allowed for the inclusion of an error correction term, which helped systematically address non-stationarity, complex short-run pricing dynamics and serial correlation, by using a dynamic OLS model (see model 2, below). Information criteria indicated an optimal lag length of one.

$$\Delta P_{it} = \alpha_i + \beta_i \Delta P_{it-1} + \gamma_i P_{it-1} + \sum_{i=1}^{i=10} \delta_i X_{it-1} + \sum_{i=1}^{i=10} \theta_i \Delta X_{it-1} + \sum_{j=1}^{j=3} \delta_{ij} D_{ijt} + \sigma_i tr D_{it} + \omega_i pw D_{it} + \varepsilon_{it} \ (2)$$

In the above model, *X* is a vector of the demand and supply variables reflected in model (1), and the dummy variables were as described above. Modelling different assumptions about the form of the transition from collusion to competition allowed us to demonstrate the impact of the assumptions on estimated overcharges. Based on model 2, above, the long-run effect of collusion on the equilibrium price was given by the ratio of the coefficient of the cartel dummy variable to that of the lagged price variable (δ_i/γ_i) [see Nieberding (2009)]. Because overcharges could be expected to have varied by product, the study allowed for the heterogeneity of both the slope and constant coefficients across the various products [see Pesaran and Smith (1995)]. This enabled direct examination and observation of differences in overcharges by product. The temporal variations in cartel overcharges were captured using dummy variables for each of cartel phases 2 and 3, as defined above.

3.5 Results

3.5.1 Regression results

The results below demonstrate that overcharges varied significantly by products and over time. For some products, the overcharges were demonstrably low and not statistically significant, while for others, the measured overcharge decreased significantly in phase 3 compared with the prior collusive phase (phase 2). This shows that the cartel's ability to exercise collective market power was imperfect, with room for competition on some products within the collusive framework (e.g. from a competitive fringe) and a responsiveness to changing market

conditions. It also supports the view that this multi-product cartel optimised its pricing on different products in line with the constraints discussed in the literature review.

3.5.1.1 Estimation results for model 2

Given the presence of non-stationarity and cointegration, the author used model 2 under three alternative transition assumptions. The results are presented in Table 3.4, below.

Table 3.4: Estimated results for model 2 under alternative transition assumptions

Model 2 under different assumptions about the transition period			
Pesaran & Smith (1995) Mean Group estimator			
All coefficients represent averages across groups (group variable: productid)			
Coefficient averages computed as unweighted means			
Mean Group type estimation			Number of obs = 570[49]
Group variable: productid			Number of groups = 10
			Obs per group: min = 57
			Obs per group: min = 57
			max = 57
ΔPrice	Model 2 with separate transition dummy	Model 2 under linear transition	Model 2 under instantaneous adjustment
	Wald chi2(9) = 3505.34 Prob > chi2 = 0.0000	Wald chi2(9) = 2922.38 Prob > chi2 = 0.0000	Wald chi2(9) = 831.71 Prob > chi2 = 0.0000
ΔPrice$_{t-1}$	-0.0735 (0.0717)	-0.0752 (0.0604)	-0.1029* (0.0609)
Price$_{t-1}$	-0.6277 *** (0.0991)	-0.6162 *** (0.0901)	-0.5754 *** (0.0920)
GDP$_{t-1}$	0.1758 ** (0.0723)	0.2034 *** (0.0722)	0.1074 * (0.0564)
Steel$_{t-1}$	0.5662 *** (0.0541)	0.5051 *** (0.0478)	0.5578 *** (0.0492)
ΔGDP$_{t-1}$	0.4206 *** (0.1166)	0.4983 *** (0.1342)	0.4224 *** (0.1207)
ΔSteel$_{t-1}$	0.5391 *** (0.0371)	0.5349 *** (0.0376)	0.5287 *** (0.0397)
ΔSteel$_{t-1}$	0.0811 (0.0604)	0.1241 ** (0.0531)	0.0548 (0.0460)
ΔZinc$_{t-1}$	-0.0247 * (0.0140)	-0.0242 * (0.0131)	-0.0311 * (0.0190)
D$_1$	0.0761 ** (0.0366)	0.0619 (0.0404)	0.0624 (0.0419)
Cartel phase 2	0.1828 *** (0.0229)	0.1624 *** (0.0220)	0.1725 *** (0.0246)
Cartel phase 3	0.1153 *** (0.0321)	0.1051 *** (0.0347)	0.1024 *** (0.0285)
Transition	0.0444 *		

---

[49] Note that this is the number of observations per variable included in the model.

	(0.0237)		
Price war	0.1030 ***	0.0871 ***	0.0982 ***
	(0.0184)	(0.0180)	(0.0219)
Constant	-1.5342 **	-1.5173 **	-1.1034 **
	(0.6403)	(0.6357)	(0.5293)

Statistically significant at: 1 per cent (p<0.01) ***, 5 per cent (p<0.05) **, 10 per cent (p<0.10) *

The results confirm the expectation that price depended on its previous values, the cost of steel, the cost of zinc, and demand. The results also confirm that, in general, collusion had a statistically significant appreciable upward effect on prices, particularly in the second and third cartel phases that were examined. This applied across all transition assumptions.

The average coefficient of the price war variable, across all products, was positive and statistically significant, indicating that while the price war might have led to a reduction in prices, it was not severe enough to result in sub-competitive pricing and was characterised by pricing above competitive levels though below the agreed cartel price path. This result is driven by the fact that in seven of the 10 products, the coefficient is positive and statistically significant, while in the remaining three products the coefficient is positive, but smaller and not statistically significant. This suggests that in the three products, the price war resulted in significantly lower prices. This is consistent with the price war not having affected all products to the same extent. Given the shorter duration of the price war compared to the overall duration of collusion and the indication that prices during the price war period were not sub-competitive, it is unlikely that consumers who received the benefit of the price war were fully compensated for the supra-competitive prices imposed by the cartel during its collusive phases.

The coefficient of the price war dummy variable is closer to the coefficient of the cartel dummy variable in the third phase of the cartel, possibly because collusion in the third phase did not adjust to the previous cartel price path because of intervention by the Commission and economic conditions which were characterised by steep cost increases.

Collusion in the third phase was less effective than in the second phase. Table 3.5, below, reflects the product-specific and cartel-phase-specific coefficients of the cartel dummy variables estimated in model 2, above.

Table 3.5: Product- and phase-specific cartel-effect coefficients

Product	Model 2 with separate transition dummy	Model 2 under linear transition	Model 2 under instantaneous adjustment
Product 1			
cartel phase 2	0.2092 **	0.1822 **	0.2149 **
cartel phase 3	0.1935 ***	0.1783 ***	0.1858 ***
Product 2			
cartel phase 2	0.0779	0.0625	0.0443
cartel phase 3	0.0089	0.0174	-0.0015
Product 3			
cartel phase 2	0.2124 ***	0.1754 **	0.2161 ***
cartel phase 3	0.0972 *	0.0830 *	0.1046 **
Product 4			
cartel phase 2	0.0795	0.0981	0.0801
cartel phase 3	-0.0625	-0.0757	-0.0643
Product 5			
cartel phase 2	0.2505 ***	0.2190 **	0.2502 ***
cartel phase 3	0.1822 **	0.1260 *	0.1812 ***
Product 6			
cartel phase 2	0.0934 **	0.0809 *	0.1067 **
cartel phase 3	0.0009	-0.0103	0.0240
Product 7			
cartel phase 2	0.2511 ***	0.2251 ***	0.2526 ***
cartel phase 3	0.1651 ***	0.1555 ***	0.1673 ***
Product 8			
cartel phase 2	0.2090 *	0.1209	0.1167
cartel phase 3	0.1410 **	0.0856	0.0780
Product 9			
cartel phase 2	0.2617 **	0.2752 ***	0.2424 **
cartel phase 3	0.1743 **	0.1910 **	0.1647 **
Product 10			
cartel phase 2	0.1833 **	0.1849 **	0.2012 **
cartel phase 3	0.2528 ***	0.3004 ***	0.1840 ***

Source: Own calculations, based on regression analyses

Table 3.6: Short-run effects of collusion on price changes under various transition assumptions

Product	Collusive phase	Model 2 with separate transition dummy	Model 2 under linear transition	Model 2 under instantaneous adjustment
Unweighted average across all products	2	20.06%***	17.64%***	18.83%***
	3	12.22%***	11.09%***	10.78%***
1	2	23.27%**	19.99%**	23.98%**
	3	21.35%***	19.52%***	20.42%***
2	2	8.10%	6.45%	4.53%
	3	0.81%	1.75%	-0.15%
3	2	23.66%***	19.17%**	24.12%***
	3	10.21%*	8.66%*	11.03%**
4	2	8.28%	10.31%	8.34%
	3	-6.06%	-7.29%	-6.23%
5	2	28.46%***	24.49%**	28.42%***
	3	19.99%**	13.43%*	19.87%***
6	2	9.79%**	8.43%*	11.27%**
	3	0.09%	-1.02%	2.43%
7	2	28.54%***	25.24%***	28.74%***
	3	17.95%***	16.82%***	18.21%***
8	2	23.24%*	12.85%	12.37%
	3	15.14%**	8.94%	8.11%
9	2	29.91%**	31.68%***	27.43%**
	3	19.04%**	21.05%**	17.90%**
10	2	20.11%**	20.30%**	22.28%**
	3	28.76%***	35.04%***	20.21%***

Source: Own calculations, based on regression analyses

In the short-run, collusion had a significant upward impact on price changes across different transition assumptions (Table 3.6, above). Products 2 and 4 do not appear to have been affected by collusion in a statistically significantly way compared with the other products. The long-run effect of collusion on the equilibrium price is given by the ratio of the coefficient of the cartel dummy variable to the coefficient of the lagged-price variable. From the results above, the unweighted average cartel effects on price across all products, as well as per product under different assumptions, on the transition are presented in Table 3.7, below.

Table 3.7: Long-run effects of collusion on equilibrium prices under various transition assumptions

Product	Collusive phase	Model 2 with separate transition dummy	Model 2 under linear transition	Model 2 under instantaneous adjustment
Unweighted average across all products	2	29.12% ***	26.36% ***	29.98% ***
	3	18.36% ***	17.06% ***	17.79% ***
1	2	17.10% **	14.65% **	18.66% **
	3	15.81% ***	14.33% ***	16.13% ***
2	2	8.23%	7.91%	5.49%
	3	0.85%	2.20%	-0.19%
3	2	108.26% ***	60.01% **	96.27% ***
	3	49.58% *	28.42% *	46.61% **
4	2	17.07%	22.25%	17.24%
	3	-13.43%	-17.16%	-13.84%
5	2	34.24% ***	32.77% **	34.32% ***
	3	24.92% **	18.85% *	24.86% ***
6	2	12.07% **	10.98% *	13.46% **
	3	-8.32%	-1.40%	3.03%
7	2	69.45% ***	52.26% ***	68.24% ***
	3	45.66% ***	36.10% ***	45.20% ***
8	2	32.85% *	22.08%	22.37%
	3	22.16% **	15.64%	14.96%
9	2	88.91% **	93.49% ***	98.38% **
	3	59.21% **	64.89% **	66.86% **
10	2	28.29% **	25.75% **	45.15% **
	3	39.02% ***	41.85% ***	41.31% ***

Source: Own calculations, based on regression analyses

There was greater consistency in the statistical significance of the long-run effect of collusion on prices across all the assumptions about the nature of the transition from collusion to competition. Product 8 was the exception, where only the model with a separate transition dummy showed that the overcharges were statistically significant. In general, collusion had a positive effect on pricing across all products in collusive phases 2 and 3, except for products 4 and 6 in phase 3, where the coefficient was negative.

The assumptions of instantaneous adjustment and of a separate transition dummy variable estimated higher overcharges than that for linear transition. In the case of instantaneous adjustment, an overestimation of the cartel overcharge was possible where the end of the cartel was characterised by sub-competitive pricing, which might have occurred if firms had entered a price war or with price cutting behaviour that was below competitive levels. This was possible given that the end of the cartel was characterised by a leniency application by one of the firms and it could have led to retaliation by the other members of the cartel. On the other hand, a

gradual decrease in the cartel overcharge towards the competitive price may have been more properly captured by a linear transition.

The effect of a separate transition dummy is to preclude the transition period from affecting the overcharge estimate. To the extent that the timing and drop in pricing between the competitive and collusive price are not accurately timed, this could have a significant impact on the estimated results. A statistically significant and separate transition dummy variable gives useful information about the impact of the end of the cartel on pricing behaviour. In this case, the transition dummy was statistically significant at the 10 per cent level of significance but showed a much smaller price coefficient compared with the phase 2 and 3 coefficients. This suggests that during this transition phase, prices had adjusted downwards compared with those during the collusive phases. The weaker statistical significance of the transition dummy variable, at 10 per cent, suggests that the pricing in this period had moved closer to the margins of competitive pricing.

The variation in the results reflects the significant impact that assumptions on the nature of the shift from collusive pricing to competitive pricing had on the results of the analyses of overcharges. The substantial variation in the overcharge estimated for product 3 across the assumptions for linear transition, on the one hand, and the instantaneous adjustment and for a separate transition dummy variable were a direct result of the relative sizes of the cartel coefficient and the coefficient of the lagged-price variable. The results show that significant consideration must be given to the assumptions made about the end of collusion and how this would affect the observed overcharges. This is particularly important if penalties are linked to the extent of the overcharges. Having different scenarios modelled allows the competition authorities to identify a range of overcharges that could have been realised by the cartel and to exercise judgement with the different ranges in mind. The determination of penalties and harm from collusion is often a matter of judgement, and the more information is available to competition authorities, the better the judgement exercised, and the better the prospects of achieving deterrence.

3.5.1.2   Heterogeneity of overcharges from multi-period collusion

This section presents a discussion on the variation in overcharges across two collusive phases – phase 2, which encompassed the period between 2001 and 2005, and phase 3, which encompassed the period between 2006 and 2008. Across all the modelled assumptions, the

cartel appeared to have been more effective in phase 2 than in phase 3. Phase 2 involved higher overcharges than phase 3, as shown in Tables 3.6 and 3.7, above. There are three potential reasons for this. The first is that phase 3 was characterised by attempts to re-establish the collusive arrangement following a period of significant instability in 2005, and it is arguable that by the time the cartel had ended, it had not yet managed to achieve the same level of efficacy as in the second phase.

The second reason for higher overcharges in phase 2 is that phase 3 coincided with sharp increases in input costs (steel prices) during 2007 – 08, in the presence of demand reduction due to the global economic recession. As such, the cartel may not have been able to pass on the full cost increases and fully exert its pricing power. The changes in demand conditions could also have resulted in a shift to a new collusive equilibrium compared with the one in phase 2. Changes in demand and supply conditions over time affect the behaviour of cartels and their ability to impose overcharges. The findings confirm that cartels are dynamic, and not static, in their response to changes in market conditions.

Finally, phase 3 coincided with increased focus on cartel enforcement by the competition authorities, and at the time, there was a focus on collusion in the steel sector. Evidence presented at the trial shows that during phase 3, some cartel members were concerned about the possibility of the cartel being detected by the competition authorities. An observation is that across all assumptions about the transition from collusion to competition and within the assumptions, there was greater variability of overcharges in phase 2 than in phase 3, despite the latter being characterised by greater input-cost volatility (see Table 3.8, below). This is consistent with observations on the variability of overcharges by product, discussed below. It might also have reflected attempts by the cartel to smooth out the impact of the volatility of input cost changes on prices, to avoid creating opportunities for cheating by members.

Table 3.8: Variability of cartel overcharges across phases 2 and 3

Phase overcharge variability (standard deviation)	Model 2 with separate transition dummy	Model 2 under linear transition	Model 2 under instantaneous adjustment	Across all assumptions
Phase 2	33.3%	25.6%	32.7%	30.0%
Phase 3	23.8%	22.4%	24.0%	22.7%

Source: Own calculations, based on regression analyses

3.5.1.3 Heterogeneity of overcharges from multi-product collusion

By allowing for heterogeneity across all specifications, the author could determine and show whether cartel overcharges varied by product, and if so, to what extent. Tables 3.6 and 3.7, above, show that there was significant heterogeneity in overcharges on different products. There were two broad clusters of overcharges, that is, products with relatively low and statistically insignificant overcharges (products 2, 4, 6 and 8[50]) and those with high and statistically significant overcharges (the rest of the products). The products falling into the lower and statistically insignificant overcharge cluster involved smoother and more stable overcharges compared with the other cluster. The variance in the estimated overcharges for the products with lower and statistically insignificant overcharges across phases 2 and 3 was 2.0 per cent compared with 7.3 per cent for those that had high and statistically significant overcharges. This could have been because where a cartel imposes higher overcharges, the incentive to cheat exists and more likely leads to greater variability in the overcharges. On the other hand, cheating for products with lower overcharges is more likely to immediately result in reversion to competitive pricing and punishment, which is sub-optimal for the cartel.

The cluster of products with lower or statistically insignificant overcharges comprised products that were considered base products, with a greater number of firms supplying them, including those outside the collusive arrangement. Steel-based products are homogenous, though there could be some differences between those products produced from scrap metal and those produced from steel billets. It is thus easier to have the cartel arrangement undercut both internally and externally on these products. Some manufacturers of these products applied for tariff protection, arguing that imports from lower-priced sources, such as those in Eastern Asia, including China, were threatening the viability of the domestic market. The products that were subject to tariff protection applications were P4, P5 and P6. The most recent tariff applications took place after the end of the conspiracy, between 2014 and 2017.[51]

The products involving greater overcharges were those that were value-added products, which reflected a further processing of the base products. These products also had fewer domestic manufacturers: primarily, the major producers involved in the conspiracy were the main sources of supply. This is consistent with the expectation that in a multi-product setting, firms

---

[50] Product 8 had a statistically significant overcharge only under the assumption of a separate transition dummy variable.

[51] http://www.itac.org.za/pages/services/tariff-investigations/investigation-reports.

are likely to set the prices of different products in line with the conditions of demand and supply pertaining to each product, to optimise the recovery of costs, for revenue and for profit maximisation. This is also consistent with pricing under cartel conduct, where firms still have an interest in optimising cost recovery and maximising cartel profits, given production costs, demand, and supply conditions.

The presence of a competitive fringe had a downward effect on prices, particularly for those products relative to the rest of the products. This does not necessarily mean pricing at competitive levels, but that the competitive fringe limits the ability of the cartel to fully exercise its collective market power. The results also support the view that in a multi-product setting, firms are more likely to collude across all products, including those for which collusion may not yield significant or higher levels of overcharges. Consistent with the observations in Chapter 2, once a multi-product firm colludes on one product, there is a possibility that it will be engaged in collusion on the other products, particularly if the same firms are involved in the production and supply of the other products. This has implications for cartel detection and deterrence.

Table 3.9, below, shows that there was significant heterogeneity in overcharges on different products during the 2005 – 06 price war period. Four of these products reflect that during the price war, the cartel did not have a statistically significant upward effect on prices. Three of the four products were those that reflected low, statistically insignificant overcharges during collusive phases 2 and 3. The fourth product had statistically significant overcharges during collusive phases 2 and 3 but had reduced and statistically insignificant overcharges during the price war. This suggests that the price war induced competition for this product, leading to a reduction in the size and statistical significance of the overcharge as firms shifted to a period of non-collusive pricing.

The cartel appeared to have been able to sustain prices significantly above the competitive level for some products, suggesting that while there may have been a price war, its effects did not include or affect all products to the same extent. In other words, the price war appears to have applied to some products, and not all products. This was expected because a price war across all products would unnecessarily undermine the overall profitability of the collusive framework, when the cheating problem could be addressed through other less costly means such as undercutting only on certain products. These findings are consistent with the evidence

from the case, which showed that for some products, the cartel continued to apply the collusive framework during the price war period.

Table 3.9: Short-term effects of price war on estimated overcharges on different products

Product	Model 2 with separate transition dummy	Model 2 under linear transition	Model 2 under instantaneous adjustment
Unweighted average across all products	10.85% ***	9.10% ***	10.32% ***
1	10.93	7.85%	12.38%
2	1.66%	0.48%	-1.15%
3	13.29% **	9.97% *	13.50% **
4	1.65%	3.48%	1.71%
5	19.33% **	16.79% **	19.29% **
6	7.54% *	6.80%	9.17% **
7	19.84% ***	16.33% **	19.91% ***
8	7.69%	1.55%	1.48%
9	14.63% *	15.62% *	13.22%
10	13.64% *	13.73% *	16.01% **

Source: Own calculations, based on regression analyses

3.6   Conclusion

The scope and number of cartels that have been uncovered in South Africa has increased significantly over the last 15 years. A sizeable number of the cartels have involved firms that produced and sold multiple products. Some of the cartels have included collusion across multiple products, across markets and using the same framework of collusion with different mechanisms, depending on how customers procured products. With more cartels being uncovered, the scope and quantum of penalties also increased significantly.

The Competition Act requires the Tribunal to consider factors such as the gravity and extent of the contravention, the loss or damage suffered because of the contravention, and the level of profit derived from the contravention when determining administrative penalties for collusion. While there may be debate about how these factors ought to be evaluated and applied, there is no doubt that firms will seek to argue that their conduct was not as harmful as alleged and did not yield significant profit. If there is sufficient information, evidence of overcharges becomes highly relevant for consideration.

It is important that all the relevant factors be considered when estimating cartel overcharges. This includes selecting an appropriate methodology based on an understanding of the dynamics

and facts of the market(s) in question, key events during the life of the cartel and the scope of the conduct. The necessary adjustments and controls must then be applied in the selected methodology. It is also important to understand the general pricing behaviour of firms involved in the cartel and how this applies in the context of the collusive arrangements under scrutiny.

Economic theory suggests that in a multi-product setting, firms may price their products to optimise the recovery of production and distribution costs, and to maximise profits, given the demand and supply conditions for each product. Where there is multi-product cartel conduct, it is likely that the effects of the conspiracy will vary by product. It is more appropriate for competition authorities, firms, and their advisors to apply methodologies that can more fully capture the dynamics of pricing in a multi-product environment and to appreciate how those dynamics apply where collusion involves multiple products where the production processes are related.

Appropriate methods in such circumstances include panel-data analyses, allowing for heterogeneity over products and time. This also means that some methodologies and assumptions are necessarily not applicable once all the facts and dynamics are considered. Examples of approaches that require caution include simple time-series analyses for individual products, but even more so those techniques that rely on simple comparisons of average prices or profits across spatial benchmarks or over time.

This chapter comprised an examination of overcharges by a multi-product cartel where there were episodes of instability in the cartel's life. In doing so, the chapter also provided an examination of the effect of assumptions about the transition period from collusion to competition on estimated overcharges, after correcting for known periods of significant instability in the life of the collusive conduct. By applying heterogeneous panel-data analyses, the study captured the cartel effect using indicator dummy variables. The unstable periods were captured by a separate dummy variable, which effectively controlled for the effects of these phases on estimated overcharges. The study found that estimated overcharges on some products were statistically significant and varied, depending on the assumptions made about the form and nature of the transition from collusion to competition. Notably, overcharges varied over products and time, despite the cartel using the same collusive mechanism throughout its lifespan.

The presence of multi-product/-market contact among firms makes collusion on several products more likely, even though the level of overcharges on some of the products may be much smaller. Base products tend to have lower, and more stable overcharges compared with value-added products.

Across all phases and within phases, the standard deviation of overcharges was greater for products with higher overcharges than for those with lower overcharges. Some of the base products had also been subject to applications for tariff protection against imports, particularly those from East Asia. This is an important insight for the competition authorities when dealing with multi-product cartels, especially given that the Competition Act requires consideration of the harm and profit derived from the infringement. This is even more relevant in an era of follow-on damages claims and innovative remedies.

The author found that overcharges varied over time and confirms that the cartel achieved a higher percentage of overcharges in its second phase than in its third phase, which was characterised by significant input-cost escalations and volatility, as well as an increased cartel enforcement focus on the steel sector. The cartel ended with the intervention of the competition authorities. The period between the second and third phases was characterised by a period of significant instability, which culminated in a price war. The beginning of the third phase was consequently characterised by efforts to re-establish adherence with the collusive arrangement. Further, the third phase was also characterised by significant escalations and volatility in the price of steel and the beginning of a significant slowdown in economic activity, which stemmed from the global economic recession. The third phase involved lower percentage overcharges than the second phase, which was more stable in terms of the collusive arrangement, input costs and economic growth. The third phase also exhibited smaller overcharge variability compared with the second phase, a finding that is consistent with findings at a product level that indicate greater variability where there are higher overcharges.

Cartel enforcement in cases where parties present claims of evidence of the overcharges achieved by multi-product cartels should include a closer examination of the methodologies applied in estimating the overcharges, considering the factual circumstances of the case at hand. For instance, attempts should be made at ensuring that any significant events in the life of the cartel – such as unstable periods and the form and nature of the transition from collusion to competition – are controlled for. This is as important as the accepted need to control for all other factors that affect prices, including demand and cost considerations. There is significant

87

risk of biased estimates if an inappropriate methodology is used and all relevant factors are not controlled for. Deterrence could be undermined if overcharges are understated.

More often, application of competition law involves significant amounts of value judgement. To ensure that judgement is applied prudently to achieve optimal deterrence in a multi-product cartel setting, it is useful to understand the effects of the cartel on the pricing of individual products and the resulting overcharges. This is more useful than merely accepting aggregate average overcharges across the cartelised products. Products with significant overcharges may be underrepresented in average overcharges. This is important because the time-series properties of aggregate data are not always akin to those of micro-units (individual products in this case). The question of how to determine appropriate overcharges and penalties in multi-product cartels is still not fully tested, at least not in South Africa.

Given that overcharges vary by product, the question is whether there is any merit in considering the factors stipulated in the Competition Act for determining penalties on a per-product basis, before examining the aggregate impact of the conduct. The standard practice has been to consider multi-product collusion as a single-cartel infringement, and it is still to be seen whether this is effective in achieving deterrence. This study proposes that overcharges should be considered and understood with respect to individual products before the competition authorities reach a conclusion on the best approach to use for determining the final penalties to impose on the conspirators. Whatever the approach, the penalties must be sufficiently high to deter firms from engaging in collusion again and to deter those firms that may be contemplating starting a multi-product cartel.

Against the background of the prevalence and features of cartels discussed in Chapter 2 and the significant size of overcharges estimated in empirical studies of overcharges – including those by a multi-product cartel such as the one studied in this chapter – in Chapter 4 the author critically examined the cartel enforcement record from the perspective of its efficacy in achieving deterrence.

APPENDIX A: PANEL UNIT-ROOT AND COINTEGRATION TESTS

Table 3.10: Probability values for unit-root tests on price, steel, zinc, other variable costs, labour and GDP

Price

Unit-root test	Null hypothesis	Statistic	p-value
Im-Pesaran-Shin (2003)	All panels contain unit roots	1.56	0.9407
Hadri LM (2000)	All panels are stationary	104.64	0.0000***

Steel

Unit-root test	Null hypothesis	Statistic	p-value
Im-Pesaran-Shin (2003)	All panels contain unit roots	2.45	0.9928
Hadri LM (2000)	All panels are stationary	109.09	0.0000***

Zinc

Unit-root test	Null hypothesis	Statistic	p-value
Im-Pesaran-Shin (2003)	All panels contain unit roots		
Hadri LM (2000)	All panels are stationary	57.68	0.0000***

Other variable costs (ovc)

Unit-root test	Null hypothesis	Statistic	p-value
Im-Pesaran-Shin (2003)	All panels contain unit roots	-12.72	0.0000***
Hadri LM (2000)	All panels are stationary	64.74	0.0000***

Labour

Unit root test	Null hypothesis	Statistic	p-value
Im-Pesaran-Shin (2003)	All panels contain unit roots	0.85	0.8014
Hadri LM (2000)	All panels are stationary	89.41	0.0000***

GDP

Unit-root test	Null hypothesis	Statistic	p-value
Im-Pesaran-Shin (2003)	All panels contain unit roots	12.98	1.0000
Hadri LM (2000)	All panels are stationary	118.04	0.0000***

*Statistically significant at: 1 per cent (p<0.01) ***, 5 per cent (p<0.05) **, 10 per cent (p<0.10) ***

The results above indicate that at least some panels contain unit roots. The study therefore tested for panel cointegration using Westerlund's (2008) panel cointegration tests. Westerland's (2008) tests allowed us to test whether some cross-sectional units were

cointegrated and whether the panel was cointegrated. To address any potential correlation between cross-sectional units, the study applied bootstrapping, which gave robust critical values. The author found that at least some of the cross-sectional units were cointegrated and that the panel was also cointegrated (see Table 3.11).

Table 3.11: Cointegration tests

xtwest lnprice lnsteel lnzinc lnovc lnlab lngdp, lags (0 2) bootstrap (1) constant				
Bootstrapping critical values under H0.				
Results for H0: no cointegration				
With 10 series and 5 covariates				
Average AIC selected lead length: 0				

Statistic	Value	Z-value	P-value	Robust P-value
Gt	-3.588	-3.126	0.001***	0.000***
Ga	-19.583	-1.768	0.039**	0.000***
Pt	-10.615	-2.878	0.002***	0.000***
Pa	-20.637	-3.538	0.000***	0.000***

*Statistically significant at: 1 per cent (p<0.01) ***, 5 per cent (p<0.05) **, 10 per cent (p<0.10) ***

Chapter 4

A critical examination of South Africa's 20-year cartel enforcement record

4.1 Introduction

Collusion has existed in South Africa for decades, taking either legal (especially in the pre-1994 era of regulation) or illegal forms. Cartels are viewed as inherently problematic arrangements that are harmful to consumers and the economy, and as such should be penalised with significant sanctions to deter firms from colluding. Some recent studies by academics and the competition authorities support this by pointing to the existence of significant overcharges by South African cartels (see Chapters 2 and 3 for indicative evidence). Consequently, there have been intensified efforts to deter the formation and continued existence of cartels, and where they form and are detected, to penalise them in a manner that ensures deterrence of the firms involved as well as those that may be considering forming similar arrangements. Admittedly, some cartels go undetected, and the size of that universe is not known. In South Africa, penalty provisions have even been revised upwards for repeat offenses to try and deter serial collusion.

South Africa has been one of the most active jurisdictions in fighting cartels, which has been enabled by several institutional arrangements. First, a leniency policy was introduced in 2004, which was later revised, in 2008. Second, the competition authorities instituted a settlement procedure, entailing that firms that co-operate and assist the Commission in prosecuting other cartel members – saving the resources of the regulator – face lower penalties. Third, since the establishment of the competition regime in its current form, greater resources – primarily through increased funding – have been committed to fighting anti-competitive conduct (using either internal capacity or external resources). Fourth, a prioritisation framework aimed at optimising the Commission's use of its limited resources for greater impact was implemented in 2006.

Although the South African competition authorities have reported on the successes of their initiatives to enhance cartel detection and prosecution, there has been no systematic study of the efficacy of the cartel enforcement record in its totality. Such a study examines the nature of the relationship between available resources, enforcement policies and tools, on the one hand, and firm prosecutions, on the other. In addition, critical empirical studies of the

interaction between leniency and settlements, on the one hand, and the duration of investigations and prosecutions relative to penalties, on the other hand, are also limited.

This chapter provides a comprehensive and critical examination of the cartel enforcement record, by looking at three areas from the perspective of deterrence. The first area examined in the study concerns the drivers of cartel enforcement in South Africa. The second area entailed examining the duration of investigation and prosecution of firms for cartel conduct. This also involved analysis of the relationship between the duration of investigations and the key drivers of cartel enforcement activity. Finally, for the purposes of understanding the potential impact of delays on deterrence, the analyses in the third area applied time-value of money principles, which are routinely used in the field of finance, to examine the impact of delays on the effective penalties paid by firms. Collectively, these three areas highlight significant gaps in cartel enforcement in its current form, and they have implications for deterrence. The three areas also highlight gaps in the economic literature on cartel enforcement. The study therefore contributes to the economic literature on cartel enforcement, policy development and efficiency improvements in cartel enforcement processes.

4.2   Review of literature on collusion and cartel enforcement

The literature review begins with section 4.2.1, which appraises the literature on the framework underlying most modern cartel enforcement regimes. The underlying framework informs not just the approach followed by competition authorities, but also the tools that they employ in cartel enforcement. The review shows that most cartel enforcement regimes operate within a deterrence-based framework, where the primary goal of the competition authorities is to deter firms from engaging in collusion. The review also addressed economic literature regarding the tools and other factors that form part of a deterrence-based framework of cartel enforcement. The study focused on the penalties, leniency and settlement procedures which are commonplace, as well as the established features of cartel enforcement across many countries.

Further to this, the literature review also provides a discussion on the empirical literature on the drivers of cartel enforcement and the effectiveness of cartel enforcement from a deterrence perspective. This includes an examination of the duration of investigations and prosecutions of firms for collusion. In turn, this entails the potential use of present-value techniques to examine the impact that delays in investigating and prosecuting firms have on the effective penalties

paid by firms, and the subsequent impact of this on deterrence. The literature review concludes by identifying some gaps in the literature, which the study attempted to fill.

4.2.1 Theoretical underpinnings of most cartel enforcement regimes

The existence of collusion and the harm associated with its conduct has led to the introduction of antitrust laws in most countries and an expansion of the tools available to competition authorities to enhance the detection, prosecution, and deterrence of cartels. This literature review covered (i) the framework governing most competition law regimes, as well as the central role played by the probability of detection and penalties (subsection 4.2.1.1); (ii) the role of leniency in cartel enforcement (subsection 4.2.1.2); and the role of settlements, including their interlinkages with penalties and leniency, and their impact on deterrence.

4.2.1.1 Deterrence-based cartel enforcement framework, probability of detection and penalties

Most competition law regimes are founded on the notion that antitrust laws must deter firms from engaging in anti-competitive conduct. This includes those that engage in collusion and are detected and punished, as well as those that may be contemplating engaging in collusion.

The economic basis of cartel enforcement derives from a substantial literature, initiated by Becker (1968). Becker developed an economics-based framework of deterrence to prevent only inefficient violations, that is, those that result in net harm to society. Becker's framework therefore allows for violations, even if these result in the transfer of a surplus from consumers to firms, provided that the net effect on total welfare is neutral or positive. The framework does not look to prohibit all violations, but instead it allows for internalisation of the costs and benefits associated with the violation. In this framework, Becker examined the costs of illegal conduct to society, in the form of costs of criminal damage (harm to society), costs to detect and prosecute criminals, and the social costs of punishment. He concluded that an optimal enforcement framework minimises these social costs.

Harm to society partly depends on the number of offences, while detection and prosecution depend on the amount of public resources dedicated to the detection and prosecution of crime, which in turn is influenced by the activities of the authority. The incentive to commit an offence (the supply of offences) is influenced by the possibility of detection and successful prosecution as well as the expected sanction. Becker (1968) found that deterrence is achieved when the risk

appetite of the offender is offset by a large enough probability of detection and prosecution as well as a fine that is sufficiently high. Becker argued that the optimal levels of the probability of detection and successful prosecution (p) and the resulting sanction (F) can be set relative to the level of net harm from the offence, the cost of raising p to a deterring level, and the effect that changes in the level of p and F will have on the supply of offences. Although there are other deterrence tools available to competition authorities, the most prevalent instrument is financial penalties. This is because they are not resource intensive, they punish criminals thereby simplifying the determination of both p and F.

Relying on the benefits and costs to society as the basis for finding the optimal level of enforcement is challenging. Such benefits and costs are difficult to quantify in most cases. For example, some of the social costs relate to the long-term damage caused by anti-competitive conduct on the functioning of markets, and as such, cannot be accurately estimated within the Becker framework. Becker's framework may therefore entail practical difficulties and may introduce enforcement inefficiencies, as much time and significant resources will be needed to quantify the costs and benefits of violations in each antitrust case. It also increases the risks of enforcement errors, as it may be difficult to distinguish between efficient and inefficient violations. The internalisation approach may also have the unintended effect of encouraging violations since firms can always claim efficiencies.

Using a deterrence-based framework for antitrust, Motta (2008) concluded that the deterrence effects of penalties depend on p and F. He found that optimal sanctions eliminate the illicit gains that offenders derive from the conduct, reducing the incentive to engage in illegal conduct. The level of the penalty must eliminate the gain obtained by offenders to reduce the incentive to engage in illegal conduct. Absolute deterrence, where p equals 1, is too costly to achieve, and optimal enforcement can be achieved by finding an appropriate balance between the level of enforcement activities and the level of sanctions [Kaplow (1989) and Bebchuk and Kaplow (1992)]. Motta (2008) suggested that if p is small, the level of sanction, F, should be set at a sufficiently high level and vice versa.

The primary question for competition authorities is therefore how to ensure that p and F are set optimally, i.e., to optimally deter firms from engaging in anti-competitive conduct, including collusion. As discussed below, optimal cartel enforcement depends, in large part, on three interrelated policies: leniency and settlement (both affecting p) and sanction (penalty), affecting F. The following paragraphs summarise the key literature related to these policies.

Bebchuk and Kaplow (1993) gave some insights on how competition authorities can achieve deterrence when enforcement activities are costly. They recognised that not every violator faces the same probability of detection. If the law enforcement agency can work out p for each violator, then the sanctions can be set optimally. However, where some violations are more difficult to detect than others, the authors suggested that the difficult-to-detect violators should face maximal sanctions, while violators that are detected more easily should face lower sanctions. In another scenario, Bebchuk and Kaplow (1993) considered a situation when the difficulty of detection cannot be observed, even after the violator has been apprehended. In such cases, the authors proposed that a single sanction should be applied.

In cartel enforcement, the competition authority rarely ever knows how easy or difficult it will be to detect collusion, and some cartels are more difficult to detect than others. As a result, the observations made by Bebchuk and Kaplow (1993) are of direct relevance to what competition authorities consider when determining penalties for cartel violations. For example, the competition authorities in South Africa routinely consider the extent of co-operation by a defendant in supplying evidence on the infringement, and whether the defendant is a repeat offender. As shown further on, monetary penalties differ across firms within and across cartels, largely because penalties are determined based on the revenue of the respective firms. There is rarely a singular monetary penalty imposed on firms for collusion, even within the same cartel, and this applies to both easy-to-detect and hard-to-detect cartels. Through penalties imposed on firms for collusion, competition authorities seek to deter penalised firms from repeating the offences, and to deter those that are contemplating collusion or that may consider it.

Block, Nold and Sidak (1981) examined the impact of antitrust enforcement on deterrence and showed that enforcement activities and credible threats of penalties affect cartel pricing by lowering it below the level that the cartel would have set in the absence of a competition authority. This price is still found to be above competitive levels, meaning that enforcement activities reduce the size of the cartel overcharge. This finding suggests that at a given level of p, penalties do not necessarily eradicate collusion but reduce the size of the optimal mark-up achieved by cartels, unless the interaction between the two reduces the cartel mark-up to negligible levels. In such circumstances, the firms will have no incentive to collude. The challenge with this finding is that most competition authorities have statutory caps on the amount of penalties that they can impose on firms and that not all cartels are detected. Given this limitation, the competition authority must then target its enforcement resources and

95

instruments in a manner that optimises *p*. This is consistent with the conclusions drawn by Becker (1968) and Motta (2008). The probability of detection at any level of cartel mark-up is influenced by the capacity of the competition authority to initiate cases, which in turn is affected by its budget. Block, Nold and Sidak (1981) showed that an increase in the United States Department of Justice's capacity and filing of complaints, reduced cartel mark-ups in the bread industry. Increased capacity of competition authorities through funding can be expected to contribute to deterrence. Bork (1978) recommended the need to increase cartel enforcement activity by adding resources, suggesting that he considered cartel enforcement to be suboptimal.

While accepting that deterrence depends on *p* and *F*, Bar-Gill and Harel (2001) argued that there are feedback effects between the crime rate and these two variables. They pointed out that *p* and *F* also depend on the crime rate and that these effects are not considered in most models. This proposal has some appeal in policymaking circles where there are debates that the level of penalties should be raised because the crime rate is high. Bar-Gill and Harel accepted the findings that a higher investment in enforcement raises *p* and lowers the crime rate but argued that this will in future have the feedback effect of reducing the size of future penalties. When future penalties are reduced, this will induce an increase in the crime rate. The policy implications of their analyses and findings are that the optimal level of investment in law enforcement depends on the effect of the crime rate on *p* and *F*. Where this effect reinforces the reduction in crime that results from investing in law enforcement and having higher sanctions, the level of investment is higher. However, if this effect reduces the direct deterrence effects of investing in law enforcement and having higher sanctions, then one can expect the optimal level of investment in law enforcement to be lower. At an analytical level, it is important to test for the existence of this bi-directional feedback effect, and where it exists, to test its strength and to apply methods that account for such effects, to avoid bias in estimates.

Landes (1983) proposed a theory for determining the optimal penalty for competition law infringements. He suggested that the optimal penalty should be equal to the net harm resulting from the conduct and that this penalty should be adjusted upwards if *p* is less than one. The optimal fine is, however, not designed to penalise efficient violations that produce net cost savings. Cartel conduct is prohibited outright regardless of whether the arrangement produces cost savings or not. These efficiency gains are applicable only in mitigation of the size of the sanction, and only in limited circumstances do they lead to a zero penalty. Posner (1976, 2001),

96

on the other hand, adopted a view that suggested the existence of under-enforcement, by advocating for higher sanctions against cartel conduct. This is an ongoing dialogue in antitrust circles with some policymakers believing that firms that engage in cartels should face greater penalties than is currently provided for in most statutes. It is true that sufficiently deterrent penalties should be capable of extracting the illicit gains earned by firms from collusion [see Motta (2008)].

4.2.1.2   Leniency, cartel detection and deterrence

The challenge for competition authorities is that, by their nature, most cartels are not readily observable to competition authorities and consumers alike. This means that competition authorities must find ways of detecting collusion. Leniency programmes have been instrumental in the detection of cartels. Section 3 presents some historical statistics on the extent to which leniency has played a role in the detection of cartels in South Africa. Motta (2008) found that leniency has played a critical role in the detection and prosecution of cartels in the EU. Despite this finding, Motta (2008) noted that leniency had not saved resources, as there was no clear reduction in the time taken to investigate and penalise firms for cartel conduct.

According to Wils (2008) the presence of a settlement procedure contributes to the prosecution of cartels once they have been uncovered. Within this context, leniency is expected to reduce the time spent investigating and prosecuting cases, freeing up resources, which can then be used to detect the existence of other cartels. Hence, if enough resources are provided to the competition authority to complement the effect of leniency, the level of p can be raised.

Harrington and Chang (2015) examined the circumstances under which leniency programmes could result in fewer cartels. The authors found that the effects of leniency on deterrence were ambiguous, depending on the impact that reliance on leniency for cartel enforcement has on the ability of the competition authorities to undertake non-leniency-based cartel enforcement. It also depends on the competition authority's budget. Leniency is found to deter collusion when non-leniency enforcement does not decline. Further, Harrington and Chang (2015) found that over-reliance on leniency in the presence of limited resources could undermine deterrence if resources shift from the detection of cartels by competition authorities to the assessment and processing of leniency applications. One of the key findings was that the instance of cartels was raised by leniency in the presence of low penalties and where processing leniency

97

applications consumed a similar amount of time as non-leniency enforcement. This might also shift resources from prosecuting active cartels to ones that are winding down. As such, deterrence is better achieved when leniency is combined with penalties that are sufficiently severe. Resource constraints can crowd out non-leniency enforcement, leading to under-deterrence especially of stable cartels. This means that the competition authorities should be well funded in order to complement leniency and not to substitute leniency with non-leniency enforcement. The problem is that the competition authorities can engage in strategic behaviour that also undermines cartel deterrence even when provided with all these tools. For instance, the competition authorities could chase after quick wins and divert resources to processing leniency applications while neglecting non-leniency enforcement. In so doing, the competition authorities may gain public and political acclaim for fighting cartels despite the negative impact on deterrence.

4.2.1.3   Settlements, time value of penalties and deterrence

Leniency is often used together with a settlement procedure in which the competition authority and a firm that has violated competition laws resolve the case by entering into a settlement agreement. Settlement procedures and their use vary across different jurisdictions. In South Africa, settlement agreements have included acknowledgement by the offending firm that its conduct has contravened the law, agreement on the future conduct of the firm and the amount of the penalty to be paid by the firm, if any. The penalty is particularly an outcome of negotiation between the Commission and the firm involved. In addition, the Commission can resolve cases using the settlement procedure at any point during the investigation, including long after it has referred cases to the Tribunal for prosecution. This has implications for the way investigations are conducted by the Commission, and the incentives and behaviour of firms under investigation when resolving cases of collusion. It also has implications for deterrence.

Settlements, like leniency, can result in benefits in terms of cartel enforcement – in the form of cost savings for both the competition authorities and the offending firms – and free up the resources of the competition authorities, which can be diverted to other prospective cases, thereby enhancing detection. However, whether a case against a firm is settled depends on several factors including the prospects of the competition authority and offending firm having converging interests in resolving case early, as well as how the surplus derived from resolving the case through settlement is shared between the competition authority and the firm [Ascione

98

and Motta (2010)]. In other words, firms expect a fair share of the settlement surplus through reduced penalties and/or less intrusive and less burdensome remedies. Edwards and Padilla (2010) found that firms have incentives to settle with the competition authorities when the sum of litigation costs and the resulting penalty are greater, and there is a high probability of successful prosecution. The incentives of the competition authorities are influenced by the likelihood of successful prosecution and the significance of litigation costs compared with the reduction in penalties if a case is resolved by settlement.

The longer the time to settlement, the more likely it will be that cost savings and other benefits of settlement to the competition authorities and hence the State will not be realised. Optimal deterrence based on settlements requires competition authorities to determine the optimal time horizon for resolving cases and when to afford firms a share of the settlement surplus through lower penalties [Franzoni (1999)]. Deterrence is undermined if the resolution of cases is prolonged and firms still obtain such a share. This is important given the potential for divergence in the incentives among firms and the competition authorities. Firms have an incentive to pay a lower real penalty when their involvement in collusion is detected, which they can achieve by delaying the penalty and still resolving the case through a settlement procedure. Holding p constant, a reduction in the penalty faced by a firm through a settlement procedure means that the required level of profit from an infringement required to offset the penalty is also lower, making it more likely that firms will commit violations in future [Ascione and Motta (2010)]. This undermines deterrence. As Franzoni (1999) observed, this means that penalties may be suboptimal for deterrence when cases are resolved using a settlement procedure. To avoid this, a reduction in penalties because of a settlement procedure should be offset by an increase in p because of increased enforcement (e.g. resources diverted to detect other cartels).

Given the simultaneous use of settlements and leniency, there is potential that significantly reduced penalties from settlements could undermine the incentives for firms to apply for leniency. Ascione and Motta (2010) suggested that this could in turn reduce the chances of detecting cartels through leniency. As such, there is a need for careful design of a settlement procedure, including the extent of fine adjustments available to firms when they resolve cases via settlements.

Katsoulacos and Ulph (2013) extended the debates on optimal sanctions to include the recognition that decisions by competition authorities may take place at different times, with

99

some being more immediate (e.g. within a year of the investigation starting), while others come after several years of investigation. They noted that time delays reduce the deterring effect of penalties, by lowering the present value of the penalty, unless the penalty is adjusted upwards. Katsoulacos and Ulph (2013) argued thus that the optimal penalty should be adjusted to account for the timing of prosecution decisions. Additional literature on optimal penalties argued for penalties to be based on cartel overcharges [e.g. Katsoulacos, Motchenkova and Ulph (2015, 2017)]. The propositions in this literature affect the size of penalties faced by firms, but do not directly address the impact of delays on the time value of penalties. Basing penalties on overcharges may create legal uncertainty and introduce implementation challenges and costs. Of relevance to the present study is the fact that it potentially prolongs the duration of cartel cases unless competition authorities and firms readily find common ground on key aspects of estimating the overcharge. Estimating overcharges is controversial and is likely to be disputed by firms because of the impact of overcharges on civil claims that firms might face after the decision of the competition authorities. If the penalties that are imposed on firms are not adjusted upwards for the potential delays that may arise from the overcharge-based approach to penalties, deterrence will be undermined further.

Edwards and Padilla (2010) argued that a populistic competition authority that places too much importance on the impact of its decisions on short-term welfare or is risk averse, or both will prefer to settle cases rather than litigate against firms, especially if the outcome of litigation is uncertain. This is likely to result in firms paying lower penalties, both nominal and real.

Based on the above review of the theoretical literature, cartel enforcement includes three important policy tools, that is, penalties, leniency, and settlements. The duration of cases from inception to a final prosecution decision is an important aspect of an optimal cartel enforcement framework because it affects the time-value of penalties. Prolonged durations of settlements result in firms paying lower real penalties, undermining deterrence.

The next section summarises the empirical literature, which covers the drivers of cartel enforcement internationally, including the use of the policy tools identified in the theoretical literature.

4.2.2 Review of literature on drivers of cartel enforcement and the effectiveness of cartel enforcement from a deterrence perspective

As the literature discussed in Chapter 2 shows, expanded cartel enforcement has contributed to the death of cartels [e.g. Levenstein and Suslow (2011)]. The availability of increased resources for cartel enforcement and a wider range of policy tools for cartel detection and prosecution have contributed to the collapse of some cartels. Some recent studies have examined the attributes of cartel enforcement over time, and how the changing use of enforcement tools by competition authorities affects cartel detection and punishment. These studies include Ghosal and Sokol (2018), who quantified the effects of various policy regimes on cartel prosecution and the financial penalties per firm and individual, and jail time in the United States. They also include an examination of the impact that political administration has on these variables. Ghosal and Sokol (2018) identified three key policy shifts in the United States cartel enforcement history, that is, the increase in the level of fines, incarceration of cartel participants and leniency. All these shifts were aimed at raising both p and F. The authors found that the introduction of a new leniency programme from 1978 to 1992 and from 1993 to 2003, the introduction of the Antitrust Amendment Act and the ACPERA period from 2004 to 2013 did not significantly influence the number of cartels prosecuted by the DoJ.

Similar conclusions were drawn on the number of individuals jailed per cartel and the numbers of individuals and firms fined per cartel. However, these policy changes significantly influenced the penalties per cartel and the number of jail days. Political party effects were significant, with a negative coefficient for the total number of cartels prosecuted during Republican administrations. Political effects were not significant for any other factors. The total number of cartels and the number of individuals jailed per cartel showed persistence, while case load, economic growth and funding did not have a significant impact. The results for the effect of funding on the number of cartels were unexpected given predictions from the deterrence-based framework of antitrust enforcement. Ghosal and Sokol (2018) reasoned that the United States Department of Justice's Antitrust Division had discretion in the allocation and utilisation of its resources across its work functions, making it difficult to formulate clear predictions. This finding is, however, consistent with studies by Ghosal (2011) and Ghosal and Sokol (2016).

The decline in the number of prosecuted cartels that Ghosal and Sokol (2018) observed could be because:

(i) The success of leniency could have resulted in fewer cartels or increased compliance.

(ii) Cartels have become smarter and more effective at avoiding detection and may be colluding tacitly via price signalling and using technology.

(iii) There has been a shift in the United States Department of Justice's focus from domestic to international cartels, and the shutdown of some field offices, together with insufficient State enforcement. The study did not include private enforcement, and the decline could simply reflect the substitution of private enforcement with public enforcement by the United States Department of Justice. Ghosal and Sokol (2018) noted that private enforcement and cartel enforcement at the level of States had not picked up local bid-rigging cartels, meaning that collusion had gone undetected. It is not possible to conclude that the observed decline was the result of effective enforcement, because the universe of cartels was not known.

Hüschelrath and Laitenberger (2015) provided an empirical assessment of the European Commission's settlement procedure, using a data set of 84 cartels penalised between 2000 and 2014. They found that the settlement procedure had led to a reduction in the time taken (of more than 12 months) between the issuance of a Statement of Objections and the time when a decision is made. This aligns with the expectation that tools such as a settlement procedure lead to resource savings and quicker decisions. However, the settlement procedure and the use of leniency do not appear to have shortened the period between the start of the investigation and the issuance of a Statement of Objections. Despite this, Hüschelrath and Laitenberger (2015) found that the settlement procedure had shortened the length of the overall period from the start of the investigation to the decision, meaning that resources can be redirected towards detecting the existence of other cartels. This could enhance deterrence.

Hüschelrath and Laitenberger (2015) recognised that the decision of a firm to settle involves more complex consideration of several factors such as the discount for settling; saving litigation costs; limiting reputational damage; the prospect of influencing the assessment of the case; and the risk of follow-on damages after admission of liability among others. It is perhaps for this reason that firms might wait for the European Commission to issue a Statement of Objections first and evaluate the strength of the case against them before engaging in more definitive settlement negotiations.

According to Hüschelrath and Laitenberger (2015), if deterrence and the quality of decisions are not undermined, the settlement procedure should enhance productive efficiency. Deterrence

can be maintained if the reduction in penalties is offset by gains in the detection rate from increased cartel enforcement activities. The study identified the potential adverse impacts on the quality of decisions, development of jurisprudence, and the impact on checks and balances, which are reinforced by appeals processes.

In summary, evidence from the empirical studies cited above provides findings that are consistent with theoretical expectations. For instance, revisions to United States leniency policy have been associated with an increased number of prosecutions per cartel, and amendments to legislation have been associated with greater penalties per cartel. The use of a settlement procedure in the European Union has resulted in resource savings, by reducing the overall duration of cases. However, this has only reduced the time from the Statement of Objections to a decision and has not appeared to affect the period from inception of investigation to the issuance of the Statement of Objectives. This suggests strategic behaviour by firms, delaying settlement until prosecution is more certain: prosecution becomes evident at the issuance of a Statement of Objectives. The next section summarises gaps in the literature reviewed.

4.2.3 Summary of gaps identified in literature

This study sought to address the following gaps that have been identified in the literature. First, the literature addresses only parts of the cartel enforcement record, rather than holistically examining the drivers of cartel enforcement and the effectiveness of enforcement from a deterrence perspective. For example, Ghosal and Sokol (2011, 2016, 2018) studied the drivers of cartel enforcement in the United States without extending their studies to evaluate the efficiency of cartel enforcement from a deterrence perspective. This study has addressed this gap by examining the drivers of cartel enforcement in South Africa; the duration of cartel cases from investigation to a final prosecution decision; and the use of present-value techniques to study the impact of duration on penalties with regard to deterrence. This is important in South Africa, where the Commission, which is well recognised for cartel enforcement, has placed much emphasis on collusion as a key enforcement priority.

The study provides, first, a critical assessment of the South African cartel enforcement record, and thus could contribute to policy changes. Second, and relatedly, to the author's knowledge, no empirical studies have examined the duration of cartel cases from investigation to the final prosecution of firms, and the relationship between duration and key drivers of cartel

enforcement activity in South Africa; this study examined these aspects of cartel enforcement in South Africa. The theoretical literature predicts that leniency and settlement procedures will reduce the amount of time to decide on cases, meaning that penalties are imposed sooner, and the resources saved could be diverted to detecting other cartels. This is expected to enhance the deterrence effects of penalties and raise the probability of detection of cartels. However, contrary to theoretical expectations, Ascione and Motta (2010) observe that leniency and settlements have not resulted in shorter durations of cases, from the start of investigations to the conclusion of cases. This finding is contrary to the more recent results of Hüschelrath and Laitenberger (2015), who found that settlements have shortened the duration of cases, allowing resources to be directed towards detecting other cartels in the EU.

Third, the unresolved aspects in the literature relate to the incentives to firms to delay the finalisation of cases, and the impact of time delays on the effects of penalties on deterrence. On a theoretical level, literature such as that of Katsoulacos and Ulph (2013) recognised that firms may benefit from prolonged cases. As such, firms may have the incentive to delay prosecution as much as possible. However, this literature does not show empirically how the incentives may work in terms of the impact on penalties paid by firms, and on the effects of penalties on deterrence. Lower real penalties undermine the effects of penalties on deterrence. This might explain the observation that the time taken from the start of an investigation to the issuance of a Statement of Objectives by the European Commission has not been reduced [see Hüschelrath and Laitenberger (2015)].

In the present study, this gap is addressed through the examination of the impact of case duration on the present-value of the penalties paid by firms at the conclusion of prosecution in South Africa. Present-value techniques are routinely used in finance to determine the time value of money and can also be applied to penalties. It can be accepted that firms know when they have participated in a cartel, hence they decide whether to resolve the case with the competition authorities soon after being notified that they are under investigation, or whether to delay resolving the case as much as possible and pay a time-discounted penalty. They may even benefit from a further discount if they can negotiate a lower nominal penalty through settlement. The incentives to firms to delay the resolution of cases will lie in the prospect of paying lower real penalties, especially where penalties are not adjusted upwards to compensate for delays. Failure to adjust penalties upwards to preserve their efficacy is likely to undermine cartel deterrence.

104

Data limitations could explain the absence of extensive empirical studies of the three areas identified above. The next section outlines the data used in this study.

4.3 Data

4.3.1 Description of data and its sources

The information used is publicly available and was obtained from the Tribunal's and the Commission's websites, annual reports, press releases and other publications. The information covered the period from April 1999 to March 2018. The study excluded information for the prior period, as the current Competition Act - which introduced more invigorated cartel enforcement – was not applicable to it. This study recognised that some cartels started prior to 1999, and some of them were legal cartels. The period prior to 1999 also coincided with State regulation of markets, where some of the key cartels were either legal (e.g. cement) or part of marketing boards. Given the variables and the period covered, the data comprises a balanced panel of six sectors (i.e., N=6) each with 19 annual periods (i.e., T=19), giving a total of 114 observations. As a general principle, having a greater number of observations provides greater comfort on the robustness of econometric analysis, particularly if the models involved are complex and require a lot of degrees of freedom. In this study, the models being estimated are parsimonious, with a limited number of variables thereby reducing the data requirements and the number of degrees of freedom required. These models are conceptually insightful in examining the relationship between cartel enforcement record (number of prosecutions and duration of prosecutions) and factors that drive them. In addition, the study does not solely rely on the results of the econometric analysis, but it is complemented with qualitative descriptive analysis. The econometric analysis provides results which are consistent with the findings of the qualitative descriptive analysis of the enforcement record data. This mitigates potential concerns about the robustness of the results of the econometric analysis.

The data relates to the completed prosecutions of firms for which there was a final ruling, by either the Tribunal or the CAC, as well as non-prosecution specific information relating to the funding of the Commission. The information on prosecution decisions includes the identity of the firms, the relevant sector or products, detail on the features of the cartel, how the case was resolved (i.e. whether settled or litigated), whether the investigation involved leniency, the date the Commission initiated the investigation, the date the Tribunal or CAC issued its judgement, and the penalty imposed on each firm for participating in collusion. It is important to note that

105

it is common for the Commission to combine multiple investigations against a firm and resolve them in a single settlement, and in this regard, there are instances where firms paid a single lumpsum penalty for multiple infringements. Also, these investigations were in some instances initiated at different points in time. In these cases, it was impossible to determine the amount of the penalty applicable to each infringement.

Information on cartel features typically includes the type of conduct (e.g. price fixing, market allocation, cover pricing, collusive tendering), when the conduct occurred, other elements of the collusive arrangement (e.g. compensation schemes) and the involvement of an industry body (e.g. industry association or third-party provider of information). The case initiation date and the Tribunal's ruling date give us a measure of the duration of the investigation from detection to prosecution. The data also comprises the number of firms that were prosecuted for collusion in each year. It is important to note that the process of prosecution can span more than a year, and this element is captured in the duration variable. Information on the Commission's funding, number of employees, case load and number of leniency applications was obtained from its published Annual Reports and relates to the Commission's financial year, not calendar year.

4.3.2 Descriptive analysis

This section starts with a brief description of the evolution of cartel enforcement since 1999 and an analysis of the evolution of the key cartel enforcement variables of interest in the study, that is, the penalties, leniency, funding and number of employees' aspects pertaining to the Commission. This is followed by a summary of the descriptive statistics comprising the key cartel enforcement variables studied.

Cartel enforcement in South Africa took its current form with the implementation of the Competition Act in 1998. The current competition authorities were set up under this Act. There was limited cartel enforcement in the early years, with the first penalties being imposed in 2003. The early cases were mostly against trade/professional associations acting on behalf of their members by setting prices or fees and against branded motor vehicle dealers (e.g. Subaru dealers, VW dealers). These cases were followed by cases involving collusion by airlines. The period after this saw more widespread cartel enforcement activity across various sectors, including agriculture and agro-processing, manufacturing, construction and transport services.

106

4.3.2.1   Evolution of penalties

The Competition Act empowers South African competition authorities to impose penalties of up to 10 per cent of revenue on firms that engage in cartel conduct. Between 1999 and 2018, the Commission collected just over R7 billion in penalties for cartel conduct. The total value of penalties received per year is influenced by the number of firms prosecuted as well as the level of penalties imposed per firm during the period. All else being equal, in periods where fewer firms are prosecuted, the total value of penalties imposed will be lower. The penalties in the earlier cases were low (see Figure 4.1). This taken together with low enforcement activity suggests that both p and F were low, and by implication there was under-deterrence. Some of the cartels that were uncovered during the first seven years were public cartels.

Figure 4.1: Evolution of penalties for cartel conduct from FY2000 to FY2018
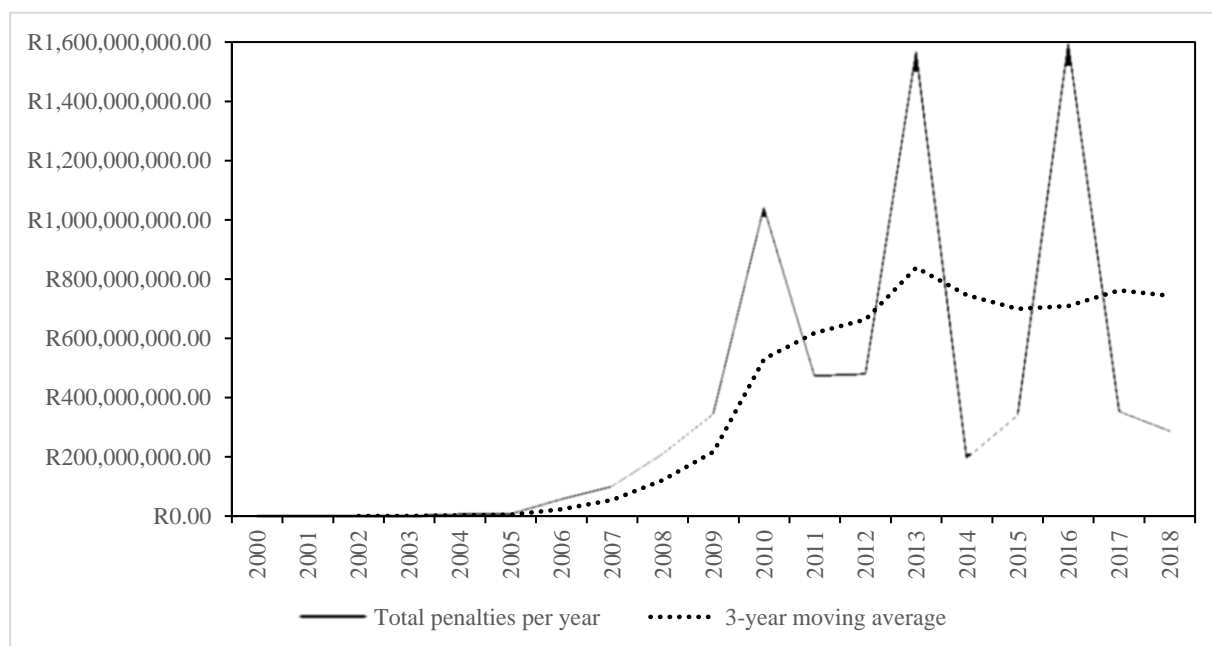
Source: Publicly available Tribunal records on penalties

Penalties for cartel conduct in South Africa started increasing from around 2005, with sharper increases from 2007. The 2007 period coincided with the detection and penalisation of the bread and milling cartels and the scrap metal cartel. Fines against Pioneer Foods for its involvement in the bread and wheat milling cartels as well as for engaging in entry deterrence strategies contributed significantly to the penalties reaching the R1 billion level in 2010. This produced the first of three outliers that are worth noting. The second peak that exceeded the 2010 level was a result of penalties against construction firms, where four of the firms that were penalised paid a total of over R1.2 billion for engaging in widespread collusion. These

107

cases formed part of the Commission's Construction Sector Fast Track Settlement Programme, an initiative developed and used by the Commission to resolve cases related to extensive collusion in the construction sector. In 2016, Arcelor Mittal agreed to pay a penalty of R1.5 billion for engaging in collusion in the steel markets, thereby producing the third peak. This was the highest single-firm penalty for collusion ever paid in South Africa. Figure 4.1 shows clearly that the general level of penalties has increased over the period but not to the level suggested by these three outliers (see three-year moving average trendline[52]).

4.3.2.2   Evolution of the use of leniency

The Commission's anti-cartel enforcement activities stepped up with the introduction of its corporate leniency policy in 2004, its revision of this in 2008, the prioritisation of sectors for more targeted enforcement, and increased public awareness following the detection of the bread cartel. The revisions to the leniency policy in 2008 gave firms more certainty, by making instigators of cartel conduct eligible for immunity, providing for the acceptance of oral evidence. Also introduced with the revisions was a marker procedure that allowed firms to submit, in the form of a marker application, a place holder for leniency while they gather evidence on their potential involvement in collusion. The marker application could then be followed by a formal application for leniency. The revisions were followed by a significant increase in the number of leniency applications. At the time, a third of the cases that the Commission referred to the Tribunal for prosecution involved leniency applications [Competition Commission (2009)].[53] Figure 4.2 presents the number of leniency applications received by the Commission over time.

---

[52] Moving averages smooth out the peaks and valleys (noise of random outliers) in the data, making it more interpretable. Moving averages also make it easier to observe the underlying general trends in data.

[53] Competition Commission, 2009, *10-Year Review, Unleashing Rivalry*, p.48.

Figure 4.2: Number of leniency applications from FY2000 to FY2018
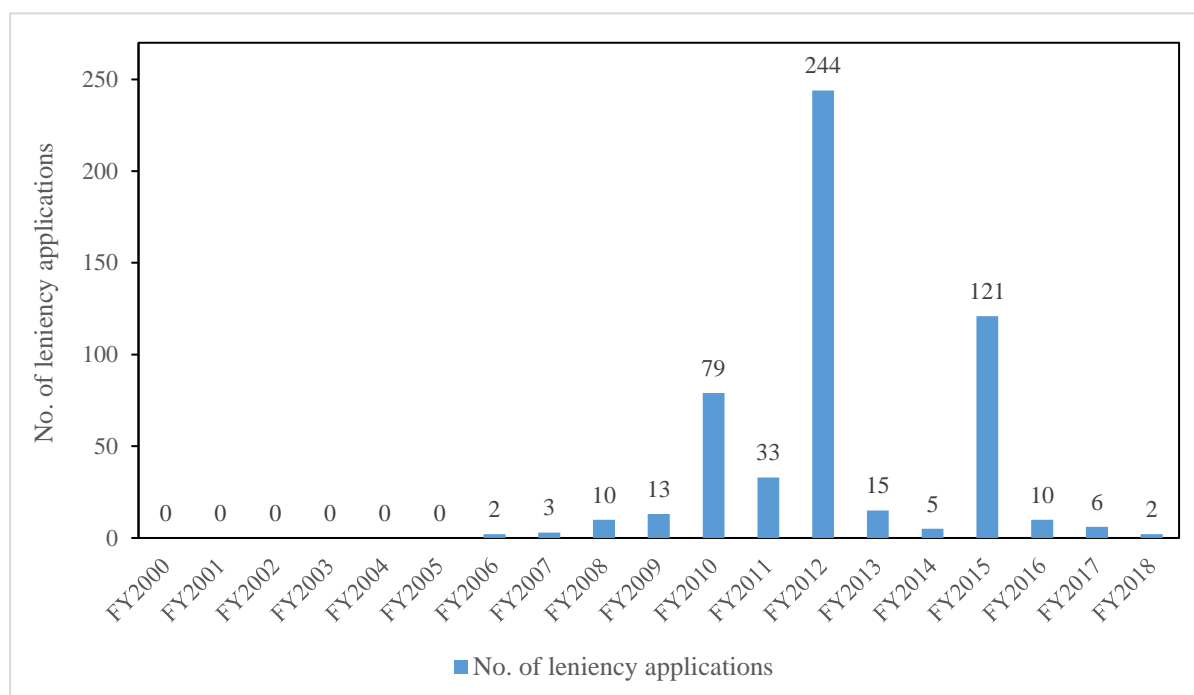
Source: Commission's Annual Reports

Figure 4.2 shows that there were three notable financial years when leniency applications increased significantly. The first of these was following revisions to the Commission's CLP and because of the construction cases referred to previously (FY2010). The second entailed the highest number of recorded leniency applications, primarily driven by cases in the construction sector following the initiation of the Commission's Construction Sector Fast Track Settlement Programme (FY2012). The third, in FY2015, was driven by applications in the automotive industry. It is important to note that these peaks in leniency applications related to cases where there was significant widespread, industry-wide collusion in the affected sectors. The Commission has stated that the key drivers of leniency applications include dawn raids in related markets, and initiation of investigations in markets where firms eventually apply for leniency [Competition Commission (2009)].[54]

The leniency policy was supported by a settlement procedure, that was expected to fast track the conclusion of investigations and prosecutions. The settlements procedure offers firms reduced penalties for the early resolution of cases, co-operation, and assistance with the prosecution of firms that opt for litigation. Over 90 per cent of the finalised cartel prosecutions during the period April 1999 to March 2018 were resolved under the settlement procedure.

---

[54] Competition Commission, 2009, *10-Year Review, Unleashing Rivalry*, p.49.

4.3.2.3 Evolution of the Commission's funding and number of employees

Under its prioritisation framework, the Commission adopted a proactive approach to initiating cartel investigations, focusing its resources and efforts on identified sectors. In 2012, the Commission set up a dedicated Cartels Division whose responsibility is to investigate allegations of cartel conduct. The evolution of the number of the Commission's employees is reflected in Figure 4.3.[55] These capacity additions in part reflect the increase in funding of the Commission between the 2000 and 2018 financial years which follows similar trends to the number of employees of the Commission (see Figure 4.3). The recent fiscal challenges facing the country are, however, putting pressure on the funding of the Commission, which increased by 8.9 per cent between the financial years 2018 and 2019, compared with previous increase of 21.8 per cent between the financial years 2017 and 2018.

Figure 4.3: Evolution of Commission's staff complement and funding[56] from FY2000 to FY2018
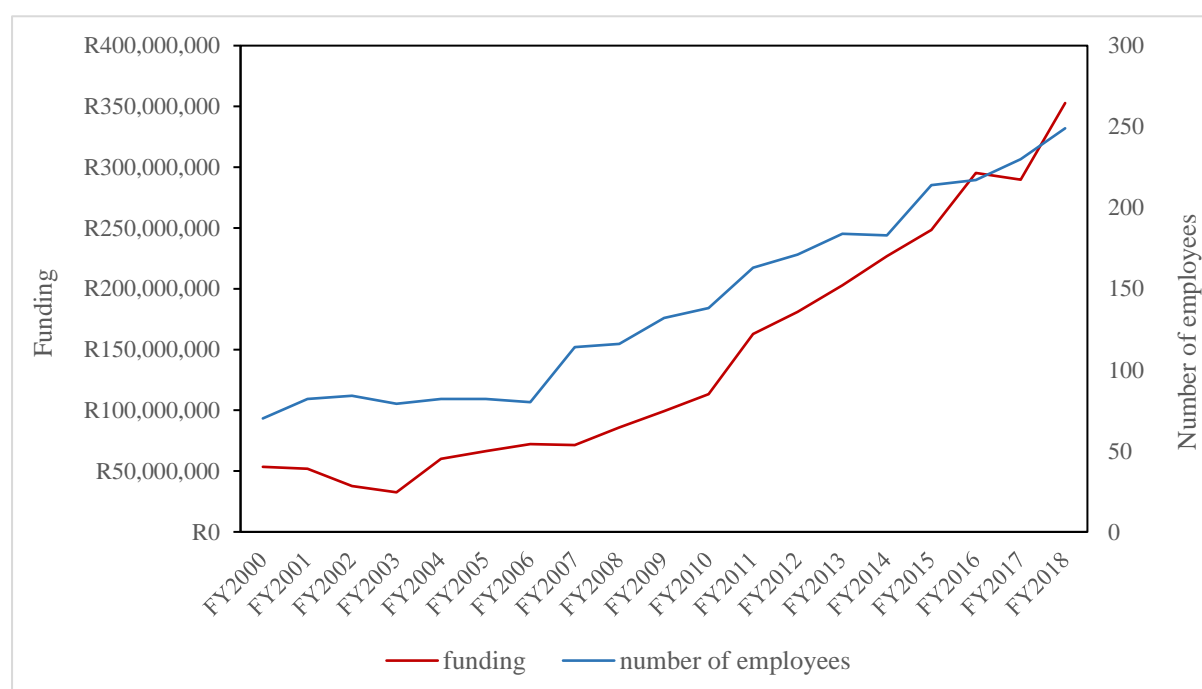
Source: Commission's Annual Reports

In more recent years, the proactive approach to cartel investigations has been complemented by intensified use of tools such as dawn raids. Between 2010 and 2018, the Commission

---

[55] These numbers include administrative and support employees.
[56] The Commission's funding includes grants from government, fees paid by firms for merger approval processes, fees for other services provided by the Commission (e.g. Advisory Opinions), and interest earned on trust funds invested.

conducted a total of 20 dawn raids, compared with four in the period between 1999 and 2009. At face value, the data shows that the combination of different enforcement tools (leniency, settlements, and proactive enforcement) and increased capacity from increased funding has raised the probability of detection and successful prosecution. It is also notable that cartel enforcement, like other branches of law enforcement, involves a significant number of learning effects, which have sharpened and enhanced the ability of the Commission to detect cartels from data, from research into markets and during routine merger review investigations. Taken together with higher penalties, these enforcement activities can be expected to have contributed to greater deterrence, at least compared with earlier years. The Commission has also applied sector-wide settlement procedures where widespread collusion was uncovered. An example of this is the Construction Sector Fast Track Settlement Programme. This can be expected to have resulted in efficiency gains in the finalisation of cases.

4.3.2.4   Summary statistics

Table 4.1 presents summary statistics for some of the key enforcement variables used in the analysis.

Table 4.1: Summary statistics on some key enforcement variables

Variable	Mean	Std. Dev
Number of firms prosecuted per calendar year	15	14
Fines per calendar year[57]	R352.1 million	R478.1 million
Fines per firm[58]	R23,7 million	R104.2 million
Number of prosecutions involving leniency per year[59]	7	9.5
Caseload per financial year of the Commission	568	156
Duration of investigation and prosecution	52.2 months (4.3yrs)	27.3
Duration of investigation and prosecution (pre-2008)	21 months (1.8yrs)	11.4
Duration of investigation and prosecution (post-2008)	54.2 months (4.5yrs)	26.9
Funding per financial year of the Commission	R142.3 million	R97.5 million
Ratio of fines to Commission's funding	1.79	2.16

Source: Own calculations based on cartel prosecution data and the Commission's Annual Reports

---

[57] Average fines per firm were affected, upwards, by the three outliers identified in Figure 1.

[58] Fines per firm were biased downwards by the low penalties in the initial years of cartel enforcement.

[59] Summary statistics for some of the variables did not provide much information given the evolution of the variables over time. For example, the average number of prosecutions involving leniency per year (leniency varied widely) and funding (which trended upwards).

The average fine per firm that is prosecuted is R23.7 million, but this varies significantly from one firm to another and from one cartel to another. Penalties depend on many factors including the revenue of the firm, the extent of co-operation provided by the firm, and other firm-specific factors that the Commission and the Tribunal may consider. There are two identifiable periods in relation to penalties: the period prior to 2007, when penalties were low compared with the period after this, and 2007 onward, when the general level of penalties imposed on firms for collusion increased.

The average duration of investigation and prosecution of cartels for the period under study was 52.2 months (4.3 years) with variations depending on the cases. Some prosecutions before the Tribunal were short, while others lasted a significant period. The Commission, as reflected in the economic literature and comparable with other jurisdictions, expected that leniency and settlements would shorten the duration of investigations and prosecutions [Competition Commission (2009), p.48]. The data on the duration of investigations and the prosecution of cartels does not confirm this.

The data for the period before the successful revisions of the leniency policy in 2008 show that the average duration of investigations and prosecution of firms was 21 months (1.8 years) compared with 54.2 months (4.5 years) post 2008. This suggests that leniency and settlements have not yielded the expected result of reducing the time to finalise investigations and prosecute firms. This could be due to several factors: first, cartels in the post-2008 period were more sophisticated compared with those in the earlier period, which were mostly non-secret cartels; second, the involvement of firms in widespread collusion (e.g. the construction and automotive sectors); and third, earlier prosecutions involved much less litigation before the Tribunal than did those in the post-2008 period. However, given that most of the prosecutions were resolved through settlements, this argument is less compelling. Settlement cases were typically resolved quickly before the Tribunal, thereby reducing the amount of time spent before the courts. Fourth, the Commission's increasing case load and number of prosecutions meant the stretching of its resources over many areas of enforcement (see Figures 4.4 and 4.5 respectively).

Figure 4.4: The Commission's average case load[60] from FY2000 to FY2018
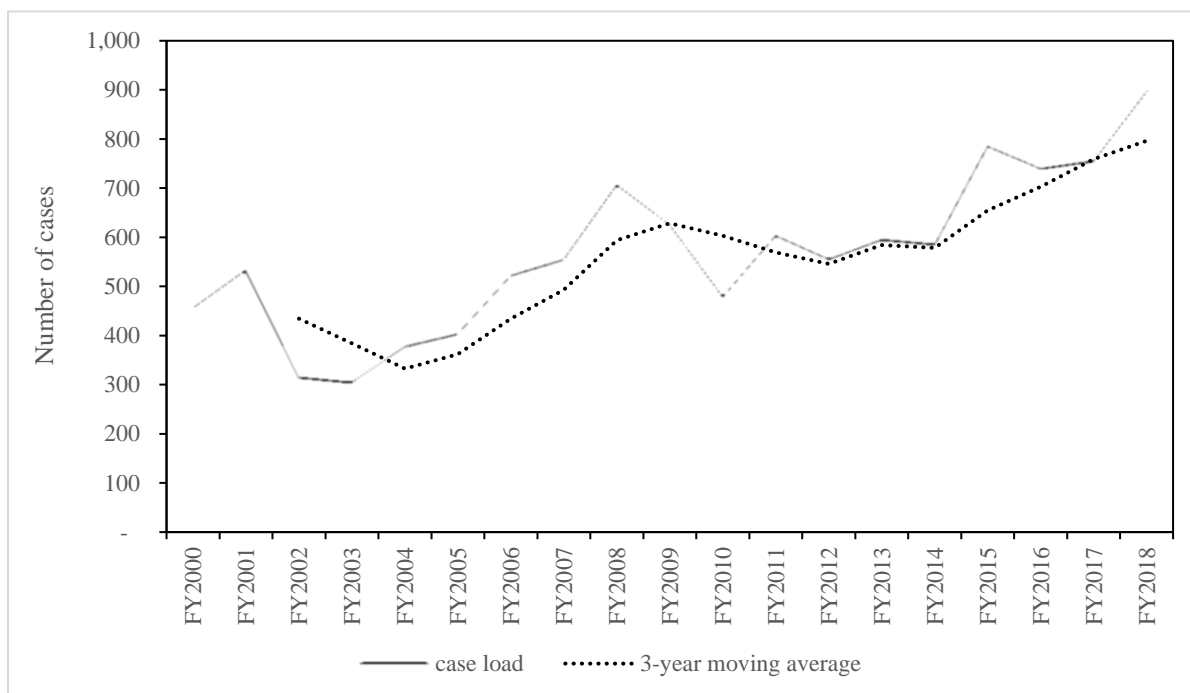
Source: Commission's Annual Reports

Figure 4.5: Number of firms prosecuted per year for cartel conduct from FY2000 to FY2018
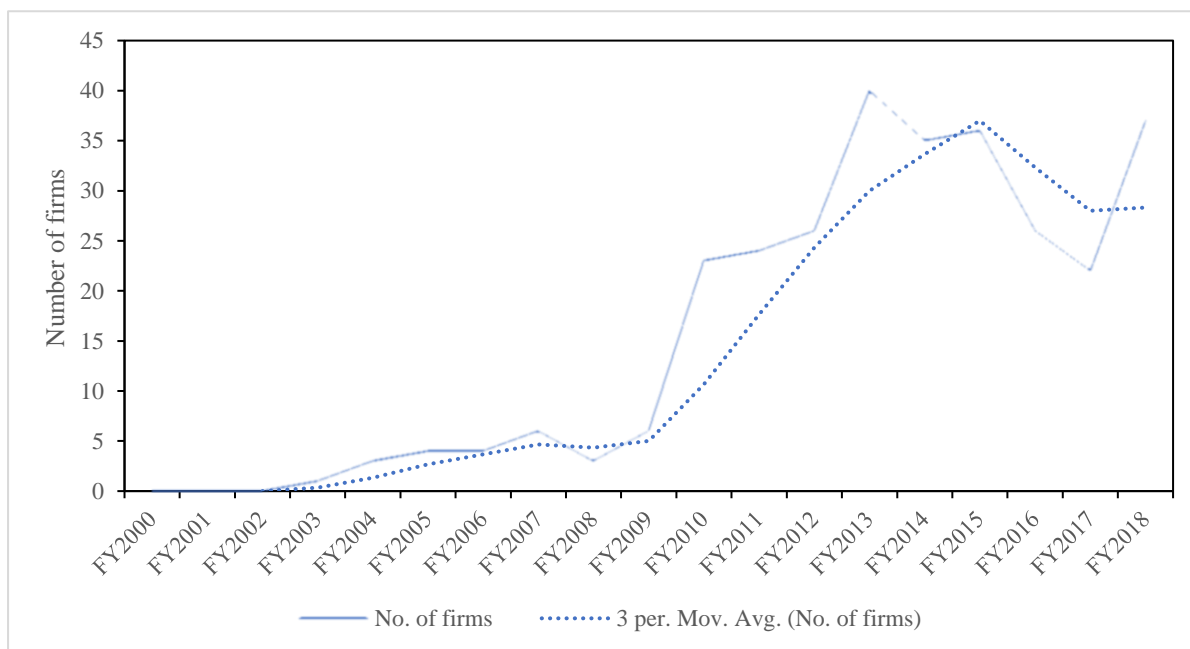
Source: Publicly available Commission and Tribunal records on prosecutions

The case load and number of prosecutions explanation has limitations because the Commission, with increased funding and number of employees, set up a standalone Cartels Division in 2011,

---

[60] Case load excludes research projects undertaken by the Commission and, more recently, market inquiries.

and the success of the leniency policy following the revisions in 2008 would have eased this challenge, unless the rate of increase in available resources, e.g. funding (Figure 4.3), had been outstripped by increases in demand for resources.[61] The increase in the number of firms prosecuted may have reflected a clearance of the backlog of existing cases or simply have been that cartels involved an increasingly higher number of firms (e.g. grain milling cartels, construction cartels, automotive parts cartels). The long duration may also be the result of internal inefficiencies within the Commission or an increased complexity of cases in the face of a reluctance by firms to supply information until they were certain that the Commission had a strong case to prosecute, and that the probability of successful prosecution was high. Observations on the duration of investigations and the prosecution of firms for cartel conduct do not diminish the importance of leniency in detecting cartels.

The cartel enforcement record summarised above shows that the South African cartel enforcement policy is founded on the principle of deterrence, and significant amounts of effort and resources have been spent trying to raise p and F to ensure deterrence. The empirical framework used in this study is discussed in the next section.

4.4   Methodology

The methodology consisted of three parts, consistent with the three research themes and described in the three subsections below. The first subsection outlines the approach used to study the relationships between the prosecution of firms, penalties, leniency and funding. The second subsection outlines the approach used to examine the duration of investigations and the prosecution of firms, given the penalties, funding, presence of leniency and a settlement procedure. The final subsection provides a discussion of present-value techniques, which provided a useful way to assess the impact of the duration of prosecution – from initiation of an investigation to final decision on the present value of penalties and its potential impact on deterrence.

---

[61] The Commission has launched several market inquiries recently. Market inquiries are complex and resource intensive.

4.4.1 Econometric methodology for examining factors driving firm prosecutions

The econometric assessment examined the impact of institutional and exogenous factors on the evolution of cartel enforcement in South Africa, particularly the prosecution of firms. Institutional factors are those factors that are within the control of the Commission, such as its internal organisation and strategies for fighting cartels. These factors include the Commission's strategic policies and tools to aid detection of collusion and prosecution of offending firms (e.g. the introduction and revision of the leniency policy and the strategic use of a settlement procedure). On the other hand, exogenous factors are those that can be characterised as being outside the control of the Commission (e.g. its funding, which is determined by the government).[62] Exogenous factors have been examined in other studies [see, for example, Ghosal and Sokol (2018)].

Other studies [e.g. Ghosal and Sokol (2018)] included GDP as an explanatory variable in the model, although its impact and statistical significance is not obvious. In all their model specifications, Ghosal and Sokol found that GDP was not statistically significant, and that result is consistent with earlier studies by Ghosal (2011) and Ghosal and Sokol (2016). This suggests that the prior empirical work does not support the inclusion of GDP. In this study, the inclusion of GDP in equation 1, generated similar results. The GDP variable may also interfere with the funding variable in that the performance of a country's economy may affect the level of funding available to competition authorities. Further discussion of the GDP variable is presented in the results section.

While case load may be expected to affect the amount of resources available to fight cartels, this is affected by the level of funding available to the competition authority. The case load variable may therefore be related to the funding variable. One may also see a rise in the case load without necessarily seeing any impact on cartel enforcement if the case load is related to other areas of enforcement (e.g. merger review and monopolisation cases). The funding variable can capture the potential impact of case load on cartel enforcement in that even if the case load increases, cartel enforcement may not suffer if funding is made available. Case load may matter where the competition authority's capacity is restricted by a shortage of funding such that it is unable to recruit more investigators or to redeploy its resources with the given

---

[62] It is accepted that the Commission submits an annual budget, although the available funding is determined by government at large. Income from fees for merger control and other services offered by the Commission cannot be controlled by the Commission. The Commission does not solicit mergers to generate income.

amount of funding. Figure 4.6, which shows the evolution of the Commission's case load and funding, shows that this is unlikely to be the case given that funding increases significantly relative to the increase in its case load, though in the latter years the Commission has also undertaken more resource-intensive market inquiries.

Figure 4.6: Evolution of the Commission's funding and case load from FY2000 to FY2018
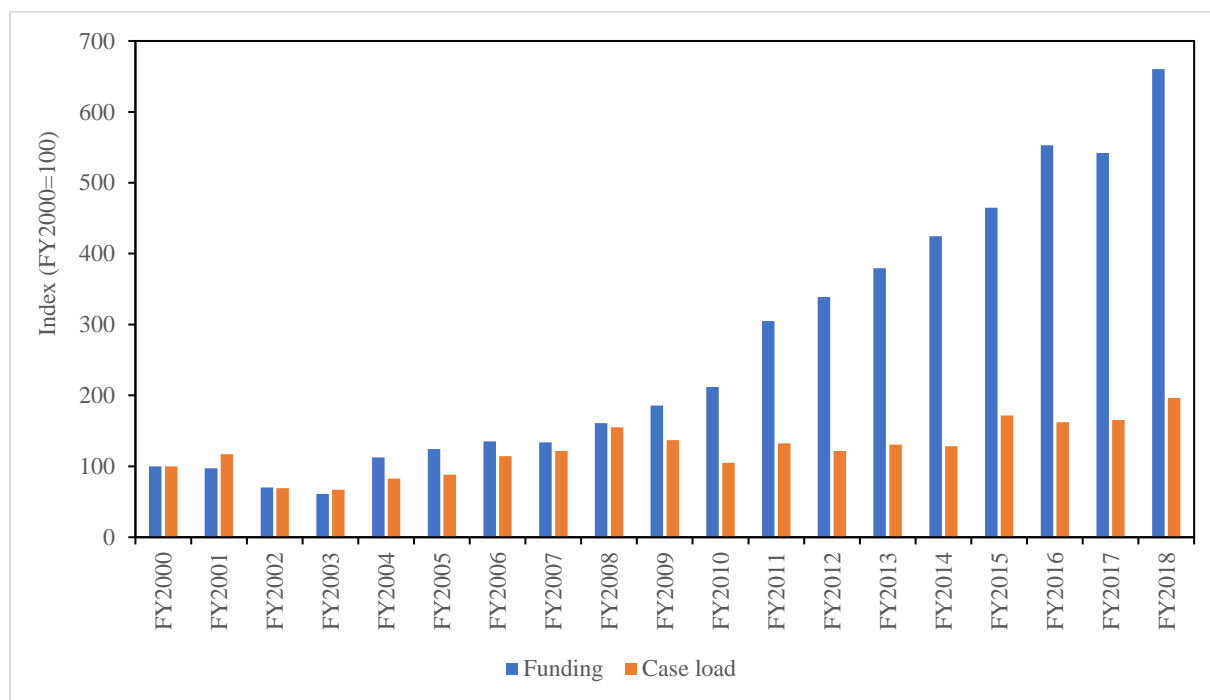
Source: Commission's Annual Reports

The inclusion of case load in Ghosal and Sokol (2018) did not find the variable to be significant across all specifications - except one - relating to the total number of days spent in jail per cartel (statistically significant only at 10 per cent). In the present study, the inclusion of case load together with funding produced results that were not statistically significant for both variables (see Table 4.12 in the Appendix for full results). Table 4.12 also presents the results for the random effects and fixed effects regressions, including case load, and the results were not statistically significant. These observations suggest that exclusion of the variable may be justified.

Within the context of a deterrence-based framework for cartel enforcement, the study was concerned with the variables that affect cartel detection, prosecution, and penalisation. These elements are important for determining the level of p and F. These variables are the amount of resources available to the Commission (captured by funding) and the policy tools available to the it (leniency, settlements, and resource deployment choices) for detecting collusion, as well

116

as the available remedies (penalties). An empirical sector-panel data model was specified to analyse the factors that influenced the number of firms prosecuted for collusion. This model was based on the panel-data model specification below.

$$firms_{it} = \propto_{0i} + \beta penalties_{it} + \delta funding_{it} + \gamma leniency_{it} + \varepsilon_{it} \quad (4.1),$$

Where:

firms refer to the natural logarithm of the number of firms prosecuted for cartel conduct in a sector per year.

penalties refer to the natural logarithm of the amounts of fines imposed on firms in a sector per year after successful prosecution. The relationship between penalties and the number of firms prosecuted for collusion can be ambiguous, in that penalties are expected to result in deterrence, meaning that over time, as more penalties are imposed, they are expected to reduce the incentive and propensity for firms to engage in collusion. However, the universe of cartels is unknown, and the more penalties are imposed on firms for collusion, the more firms may confess to their involvement in collusion, leading to greater prosecutions of firms.

funding refers to the natural logarithm of the size of financial resources available to the Commission per year for its use. A positive relationship between funding and the number of firms prosecuted for cartel conduct can be expected since the greater the resources available to the Commission, the greater the ability to detect, investigate and prosecute firms. This is the case if the resources are used efficiently and effectively.

leniency refers to the natural logarithm of the number of cartel prosecutions involving leniency per sector per year. Leniency was expected to enhance the detection of cartels, and as such, increased the number of firms prosecuted for collusion. As a result, a positive relationship can be expected between leniency and the number of firms prosecuted for involvement in cartel conduct.

It was not possible to determine the amount of funding the Commission allocated to each sector in the period of study, even in the presence of a prioritisation framework: the publicly available annual report of the Commission did not have this breakdown. The Commission investigates all complaints across all sectors and will allocate resources to all sectors. In doing so the Commission has some level of discretion over how it deploys and balances the use of its

resources given its responsibilities, goals, and strategic priorities. For example, if it prioritises cartel enforcement, it could decide to dedicate more resources (funding and manpower) towards that enforcement. This, of course, has implications for the Commission's ability to pursue other enforcement priorities such as advocacy, merger reviews, pursuance of restrictive business practices, abuse of dominance, market inquiries, and general market research and monitoring. For estimation purposes, the author assumed that an increase in the Commission's funding would enhance its ability to investigate and prosecute more firms across all sectors where collusion had been detected.

4.4.2 Approach to examining duration of investigation and prosecution of firms

The examination of the duration of investigation and the prosecution of firms was twofold. First, a simple qualitative examination was made of the time taken from initiation of a case to the decision of the Tribunal and/or CAC. Trends in the duration of cases were subsequently examined in the study given the use of leniency, settlements, penalties, and the funding of the Commission. Theory predicts that leniency and a settlement procedure, taken together with an increase in available resources, should reduce the amount of time taken to investigate and prosecute firms for collusion.

A second aspect of the examination of the duration of investigations and prosecutions of firms entailed the use of panel-data analyses, with duration as the dependent variable and penalties, funding, leniency, and number of firms as explanatory variables (see equation 4.2, below). Leniency and increased capacity through increased funding were expected to have a negative relationship with duration. The relationship between duration and penalties can be ambiguous. On the one hand, the imposition of higher penalties over time can be expected to encourage firms to come forward and apply for leniency and/or settle cases earlier with the competition authority. These firms would also provide valuable evidence on the cartel(s), which could be used to prosecute other firms. This could be expected to reduce the duration of investigating and prosecuting firms. However, if firms have incentives to engage in strategic behaviour, which delays investigations and penalties, and there is no upward adjustment of penalties for delays, a positive relationship between penalties and duration can be observed. The model also included the number of firms prosecuted for collusion, as investigating and prosecuting more firms can be expected to increase a competition authority's workload and potentially increase case duration. Since duration is trended (see Figure 4.7, below), the model included a trend variable.

$$duration_{it} = \propto_{0i} + \beta penalties_{it} + \delta funding_{it} + \gamma leniency_{it} + \rho firms_{it} + \theta trend + \varepsilon_{it} \text{ (4.2),}$$

where:

duration refers to the natural logarithm of the average length of time taken to investigate and prosecute firms for collusion per year per sector.

firms, penalties, funding, and leniency are as defined above.

trend: because the duration variable is trend stationary, the model included a trend variable.

The results of the qualitative and quantitative analyses are compared for consistency.

4.4.3 *Approach to assessing the impact of delays on penalties*

To assess the impact of delays on penalties, the author employed the present-value technique, which is rooted in the finance concept that money received today is worth more than the same amount received at a point in the future. The same applies to the impact of penalties on a firm's decision to collude in the presence of the risk of a penalty. A R100,000 penalty paid earlier has a greater deterrent-effect on firms than a R100,000 penalty paid five years after the cartel was first detected. The value today of a R100,000 penalty paid in 5 years' time, at a discount rate of 10 per cent is approximately R60,657[63], assuming daily compounding. This is a massive discount on the penalty, which results purely from delaying payment of the R100,000 penalty by five years.

In the field of finance, the value of money is preserved by ensuring that interest is either charged or earned over time. In cartel enforcement, a similar concept is rarely applied. Competition authorities rarely include interest to preserve the deterrence effect of penalties when firms delay the finalising of cases, especially where cases are eventually settled. A firm's decision on whether to co-operate with the competition authority and when to negotiate with the competition authority is based on the value it derives from either early or delayed resolution. This decision is therefore affected by policymakers choosing not to account for interest.

A penalty paid by a firm when a case is resolved through settlement is an outcome of negotiation between the Commission and the firm. Firms have every incentive to negotiate

---

[63] The discount rate and the frequency of compounding have a significant impact on the present value of an amount paid in the future.

paying no penalty at all or paying the lowest possible penalty they can pay, even after accepting that they broke the law. A cartel member derives greater utility from delaying payment of the penalty that it expects or is prepared to pay because this effectively amounts to paying a lower penalty than if the penalty was paid earlier [see for example, Katsoulacos and Ulph (2013)]. This is the case especially if firms face the prospect of significant penalties, unless there are other considerations that offset the benefit of delay. An example of this would be the firm suffering other forms of harm from having a prolonged investigation or prosecution, such as continued public regulatory scrutiny and reputational harm. If these further costs to the firm – taken together with the penalty that it ultimately pays if the firm is successfully prosecuted – are greater than the benefit of delay, then the firm has a greater incentive to expedite finalisation of the case by co-operating, concluding settlement agreements with the Commission and paying the penalty.

The cartel case data in South Africa shows that the duration of investigations and the prosecution of firms involved in cartels has been rising over time, meaning that firms are increasingly taking longer to finalise settlement agreements with the Commission. Similarly, the data shows that most cartel prosecutions are concluded through settlement between firms and the Commission. Wils (2008) argued that for the purposes of ensuring optimal deterrence, settlements should not result in a situation where enforcement losses from settling cases with firms (e.g. lower penalties) outweigh the expected enforcement gains derived from quicker resolution of cases (e.g. lower costs, shorter investigations and opportunities to redirect resources to detect other cartels). The South African enforcement record suggests that there may be too much reliance on settlement and leniency, and this may affect deterrence, as noted by Edwards and Padilla (2010). Delays in the finalisation of cases mean that the expected benefits of having a leniency procedure complemented by a settlement procedure are not realised.

The present-value of a lump-sum amount (penalty) paid by a firm in the future can be calculated using the formula in equation 4.3, below:

$$PV_{penalty} = \frac{F}{(1+r/_{nT})^{nT}} \quad (4.3),$$

where:

PV$_{penalty}$ is the present-value of the penalty for a firm that is successfully investigated and prosecuted for collusion.

F is the penalty paid by the firm after successful investigation and prosecution.

r is the discount rate, represented by the average yield rates on five- to 10-year government bonds during the respective period between the start of the investigation and finalisation of prosecution of the firm. The author uses five- to 10-year discount rates because a significant number of investigations and prosecutions lasted for over three years, and most were settled within a period of four to 10 years. Government bonds entail a lower risk compared with other types of investments.

n is the number of times the discount (interest) rate is compounded per year.

T is the number of years between the start of the investigation and penalisation. Commencement of an investigation increases the likelihood of cartel detection and punishment of firms, and under the current framework, firms can settle cases with the Commission at any stage between the start of the investigation and the finalisation of prosecution.

4.5 Empirical results

The empirical results are presented in the same order as the discussion in the methodology section, starting with the econometric results for the relationship between the prosecution of firms, on the one hand, and penalties, leniency and funding, on the other. This is followed by an examination of the duration of investigation and prosecution of firms and, lastly, an examination of the impact of duration of investigation and prosecution on the present value of penalties, and its potential impact on deterrence.

4.5.1 Results of econometric analysis of factors driving firm prosecutions

4.5.1.1 Unit-root tests

Before presenting the results of the panel-regression analysis, the data was tested for the presence of unit-roots to ensure that the results were not spurious. The presence and impact of persistence shocks affects the validity of regression results, and unit-root tests help to detect the presence of persistent shocks, which should be accounted for when estimating regressions.

T-ratios do not follow a *t*-distribution in the presence of a unit-root, meaning that hypotheses about the regression parameters cannot be validly tested. Therefore, author tested the following variables for unit roots: the number of firms prosecuted per year per sector; the amount of penalties imposed on successfully investigated and prosecuted firms per year per sector; the funding of the Commission; and institutional factors, which included the number of prosecutions that involved leniency. As mentioned previously, the data comprised a balanced panel of six sectors (i.e., N=6), 19 annual periods (i.e., T=19) and four variables.

Several unit-root tests are available for panel data, each with particular assumptions, strengths, and limitations. The Levin-Lin-Chu (2002) unit-root test assumes a common auto-regressive parameter for all panels, but it does not allow for the possibility that some sectors have unit-roots. It also requires that the ratio of the number of panels (N) to the number of time periods (T) tends towards zero as T grows. To mitigate the effects of cross-sectional correlation, cross-sectional means were subtracted. The unit-root test developed by Harris and Tzavalis (1999) can be applied to panel data sets with a small T – by allowing for small sample adjustments to T – and assumes, unlike the Levin-Lin-Chu test, that there is a homogenous variance. The test can be applied with balanced panel data.

In the Breitung (2000) test, the data is adjusted prior to fitting a regression, meaning that it is not necessary to undertake bias adjustments. This is contrary to the Levin-Lin-Chu unit-root test. The Breitung test has much higher power than the Levin-Lin-Chu test when autoregressive parameters approach one and when panel-specific effects are included. The Breitung test also has good power in small datasets. However, the power of the test falls when T is fixed and N increases. This was not a problem, though, with the dataset used in the study. The test was also robust to cross-sectional correlation. The Im-Pesaran-Shin (2003) test relaxes the strict assumptions of the Levin-Lin-Chu test that all panels have a unit-root, by allowing for only some panels to have a unit root. The disadvantage of this test is that its power reduces significantly if a substantial number of the panels have unit roots. In panel-data analysis, one can also use the Fisher-type [Choi (2001)] unit-root test, which combines the p-values from unit-root tests on individual panels to obtain an overall test statistic. This enabled the author to determine whether the panel series had a unit root.

This study relied on all the unit-root tests described above, which allowed the examination to take advantage of the strengths of the different tests. In turn, this gave greater confidence in the validity of the results on the unit-root tests and the panel-regression results. These tests have

the null hypothesis that all or some of the panels have a unit root. The tests showed that the variables contained in our six panels of interest were stationary, since in all instances, the null hypothesis that the panels have unit roots was rejected. Notably, the funding variable was trend stationary. The existence of a trend in this variable is also clear in Figure 4.3, above. Table 4.2, below, presents the results of our unit-root tests. Given these results, the study proceeded to estimate the panel regressions from the data, in levels.

Table 4.2: Panel unit-root test results

Test	Variable	Statistic	p-value
			Number of panels = 6
			Number of periods = 19
H_O: panels contain unit-root			
Levin-Lin-Chu	Firms	-2.5165	0.0059***
Harris-Tzavalis	Firms	-4.3651	0.0000***
Breitung	Firms	-4.9729	0.0000***
Im-Pesaran-Shin	Firms	-3.6219	0.0001***
		P = 52.9378	0.0000***
		Z = -5.0193	0.0000***
Fisher-type	Firms	L* = -5.9513	0.0000***
		Pm = 8.3564	0.0000***
Levin-Lin-Chu	Penalties	-2.2891	0.0110**
Harris-Tzavalis	Penalties	-4.6083	0.0000***
Breitung	Penalties	-3.4765	0.0003***
Im-Pesaran-Shin	Penalties	-3.3755	0.0004***
		P = 67.8491	0.0000***
		Z = -5.1623	0.0000***
Fisher-type	Penalties	L* = -7.5362	0.0000***
		Pm = 11.4002	0.0000***
Levin-Lin-Chu	Funding	-10.0220	0.0000***
Harris-Tzavalis	Funding	-3.1108	0.0009***
Breitung	Funding	-1.6717	0.0473**
Im-Pesaran-Shin	Funding	-4.2524	0.0000***
		P = 40.9460	0.0001***
		Z = -4.5042	0.0000***
Fisher-type	Funding	L* = -4.6335	0.0000***
		Pm = 5.9086	0.0000***
Levin-Lin-Chu	Leniency	-2.2797	0.0113**
Harris-Tzavalis	Leniency	-2.6724	0.0038***
Breitung	Leniency	-3.3768	0.0004***
Im-Pesaran-Shin	Leniency	-2.3383	0.0097***
		P = 36.4573	0.0003***
		Z = -3.0806	0.0010***
Fisher-type	Leniency	L* = -3.7164	0.0004***
		Pm = 4.9923	0.0000***

*Statistically significant at: 1 per cent (p<0.01) ***, 5 per cent (p<0.05) **, 10 per cent (p<0.10) ***

4.5.1.2   Panel regression analyses

In the panel regression analysis, the author started by testing whether the appropriate model was the fixed-effects model or the random-effects model. This was decided based on the Hausman panel-specification test, which has the null hypothesis (H_o) that the random-effects model was appropriate and an alternative hypothesis (H_1) that the fixed-effects model was appropriate. Table 4.3, below, shows the results of the test.

Table 4.3: Hausman model specification test results

Variables	Coefficients		(b - B)	Sqrt (diag (V_b-V_B))
	(b)	**(B)**	**Difference**	**S.E.**
Firms	fe	Re		
Penalties	0.0727181	0.0744654	-0.0017473	0.0020461
Funding	0.0695592	0.0678405	0.0017187	0.0146004
Leniency	0.3859787	0.3661721	0.0198066	0.0262031

b = consistent under H_o and H_a; obtained from xtreg
B = inconsistent under H_a, efficient under H_o; obtained from xtreg

Test: H_o: difference in coefficients not systematic
    χ^2 (3) = (b-B)' [(V_b-V_B) ^ (-1)] (b-B) = 0.89
    Prob > chi2 = 0.8281

The test statistics show that the null hypothesis – that the random-effects model was the more appropriate model to use – could not be rejected. Nonetheless, the results for both the random-effects and fixed-effects models are presented. Panel regressions were estimated at an aggregate level using the bootstrapping technique, which has the advantage of allowing us to control and check the stability of the regression results. Bootstrapping also produced asymptotically more accurate results than the standard intervals produced using sample variance and assumptions of normality. The results of the random-effects and fixed-effects regressions are presented, respectively, in Tables 4.4 and 4.5, below.

Table 4.4: Results of random-effects panel regression analysis of the impact of penalties, funding, and leniency on the number of firms prosecuted for collusion in South Africa

xtreg: *number of firms prosecuted = f (penalties imposed, funding of the Commission, use of leniency to detect collusion)*

Bootstrap replications (50)

Random-effects GLS regression		Number of observations = 114
Group variable: sector		Number of groups = 6
R-sq: within = 0.8860		Obs per group: min = 19
between = 0.9616		average = 19.0
overall = 0.8949		max = 19
		Wald Chi2(3) = 451.27
corr (u_i, x) = 0 (assumed)		Prob > Chi2 = 0.0000
		(Replications based on 6 clusters in sector)

Firms	Observed coefficient	Bootstrap Std. Err.	Z	P > |z|	Normal-based [95per cent conf. interval]	
Penalties	0.0744654	0.0078369	9.50***	0.000	0.0591055	0.0898253
Funding	0.0678405	0.0362308	1.87*	0.061	-0.003171	0.1388516
Leniency	0.3661721	0.0978371	3.74***	0.000	0.174415	0.5579292
Constant	-1.228993	0.6552114	-1.88*	0.061	-2.513183	0.0551982
sigma_u	0.03438046					
sigma_e	0.30381385					
Rho	0.01264393 (fraction of variance due to u_i)					

*Statistically significant at: 1 per cent (p<0.01) ***, 5 per cent (p<0.05) **, 10 per cent (p<0.10) ***

Given that equation 4.1 was specified in log-log form, the coefficients of the explanatory variables are elasticities measuring the per cent change in the dependent variable when the respective explanatory variable changed by 1 per cent. The results of the random effects regression estimated above can therefore be interpreted as follows: a 1 per cent increase in penalties increased the number of firms prosecuted for collusion by 0.07 per cent; a 1 per cent increase in funding for the Commission increased the number of firms prosecuted by 0.07 per cent; and a 1 per cent increase in leniency applications increased the number of firms prosecuted for collusion by 0.37 per cent. Stated differently, a 10 per cent increase in the number of leniency applications led to an increase in the number of firms prosecuted for collusion by 3.7per cent. The impact of leniency on the detection of those penalised cartels was likely higher. The penalties and leniency variables were statistically significant at the 1 per cent level, while the funding variable was weaker and statistically significant at the 10 per cent level.

This may have been a result of the limitations of the funding variable, as discussed in the methodology section.

The impact of penalties on the number of prosecutions shows that penalties increased the likelihood that more firms would be prosecuted, but the effect was much smaller. A 10 per cent increase in the level of penalties resulted in an increase of only 0.7 per cent in the number of firms prosecuted for collusion. Given that most cases were resolved through leniency and settlements, penalties can be seen to have incentivised firms to come forward, but the level of penalties may not have been severe enough to have a large impact on the number of prosecutions. Of these variables, leniency was the main driver of prosecutions, which is consistent with the qualitative analysis, which showed that over 90 per cent of cases involved leniency.

Table 4.5: Results of fixed-effects panel regression analysis of the impact of penalties, funding, and leniency on the number of firms prosecuted for collusion in South Africa

xtreg: *number of firms prosecuted = f (penalties imposed, funding of the Commission, use of leniency to detect collusion)*

Bootstrap replications (50)

Fixed-effects (within) regression				Number of observations = 114	
Group variable: sector				Number of groups = 6	
R-sq: within = 0.8861				Obs per group: min = 19	
between = 0.9589				average = 19.0	
overall = 0.8947				max = 19	
				Wald Chi2(3) = 736.38	
corr (u_i, xb) = 0.0093				Prob > Chi2 = 0.0000	
				(Replications based on 6 clusters in sector)	

Firms	Observed coefficient	Bootstrap Std. Err.	Z	P > |z|	Normal-based [95per cent conf. interval]	
Penalties	0.0727181	0.0097466	7.46***	0.000	0.0536152	0.091821
Funding	0.0695592	0.0385580	1.80*	0.071	-0.006013	0.1451315
Leniency	0.3859787	0.0969151	3.98***	0.000	0.1960287	0.5759288
Constant	-1.257367	0.6935727	-1.81*	0.070	-2.616745	0.1020103
sigma_u	0.07029178					
sigma_e	0.30381385					
Rho	0.05080977 (fraction of variance due to u_i)					

*Statistically significant at: 1 per cent (p<0.01) ***; 5 per cent (p<0.05) **; 10 per cent (p<0.10) **

The interpretation of the fixed effects model results is the same as that of the random effects model in that the coefficients are elasticities and can be interpreted in the same way. The results for the fixed-effects estimation were consistent with those for the random effects estimation and showed that all these variables have a positive relationship with the number of firms prosecuted for collusion. Penalties and leniency were statistically significant at the 1 per cent level of significance, while funding was statistically significant at the 10 per cent level of significance. These results are consistent with the results of the descriptive analysis presented in Figures 4.1, 4.2 and 4.3, read together with Figure 4.5. It is widely accepted that leniency has been an important tool for cartel enforcement in terms of detection and prosecution in South Africa. A significant number of cartel prosecutions were finalised by way of a settlement procedure and in many instances involved leniency. The results on leniency are significant given the positive impact it has on cartel enforcement. It will likely remain key to future cartel enforcement by the Commission given the importance of expanding the scale of cartel enforcement in order to enhance deterrence.

The Commission has received increasing levels of funding on a year-to-year basis, which has enhanced its capacity to investigate and prosecute cartel conduct. The availability of resources has enabled the Commission to set up a standalone Cartels Division, whose resources are devoted to, and focus is solely on detecting and investigating cartel conduct. In general, despite the limitations of the funding variable, the results support the hypothesis that increasing the level of funding for the Commission, as well as enhancing the leniency policy would contribute to increased cartel detection, investigation, and prosecution. This, however, gives limited insight into how effectively the increased funding over time and the leniency policy have been deployed to ensure administrative efficiency and effective deterrence. This author attempted to answer these questions, as shown in the sections that follow.

Per the methodology discussion, above, the inclusion of GDP in both the random-effects and fixed-effects regressions produced results in which penalties and leniency remained statistically significant at the 1 per cent level of significance, but GDP and funding were not statistically significant. The coefficient was -0.028 suggesting that a 1 per cent increase in GDP would have reduced the number of firms prosecuted for collusion by 0.028 per cent.

To address potential concerns about endogeneity between the number of firms prosecuted and leniency applications, for example, the Blundell and Bond (1998) systems generalised method of moments (GMM) estimator was used to estimate a dynamic panel data model. The results

of the systems GMM estimation are presented in Table 4.9 in the Appendix. The results confirm that leniency and penalties are positively related to the number of firms prosecuted and generate coefficients that are of similar magnitude and level of statistical significance. The coefficient for leniency is 0.3612 under the systems GMM estimation while it is 0.3662 and 0.3860 under the standard random-effects and fixed-effects estimations, respectively. These results confirm the importance of leniency as a key driver of cartel detection and prosecution. Similarly, the coefficient for penalties is 0.0762 under the systems GMM estimation while it is 0.0745 and 0.0727 under the standard random-effects and fixed-effects estimations. The difference between the systems GMM estimation results and the results of the random-effects and fixed effects estimations in Tables 4.4 and 4.5 is that the coefficient for funding, though having the same sign, differs significantly and is not statistically significant.

4.5.2 Examination of duration of investigation and prosecution of firms

One of the expected contributions of leniency and increased capacity to detect and prosecute cartels is that they would have enabled the Commission to deal expeditiously with cartel investigations, reducing the duration of time from initiation of investigation to penalisation. This is indeed what economics theory envisages, and it is one of the key motivations for leniency policies in competition law enforcement. That is, leniency is expected to adduce evidence, which is then used to prosecute other cartel members, thereby speedily freeing up resources, which can then be directed toward detecting those cartels that the competition authority had not yet uncovered. Leniency can be enhanced by the presence of a settlements procedure that allows the Commission to resolve cases with firms and to agree on, amongst other things, the amount of penalties to be paid.

The question that then arises is whether the positive relationship observed between leniency and prosecution has resulted in all the envisaged benefits that economic theory anticipates or whether firms have exploited it to reduce the effective penalties, without necessarily resulting in investigative and prosecutorial efficiency. In the discussion below, the study revealed that leniency has not necessarily led to a reduction in the duration of investigations and the speedy prosecution of firms. This is reflected in Figure 4.7, below, which shows that the average duration of investigations and prosecutions has been increasing.

There are several potential reasons for the increasing duration of cases. First, cartel cases have become more complex e.g. the existence of endemic, industry-wide collusion, requiring

complex arrangements to resolve (e.g. the construction sector collusion). Second, there is inefficiency on the part of the Commission in resolving cases, even where leniency exists. For instance, some of the cases resolved by the Commission in 2018 were from the 2006/07 to 2007/08 period. Among these cases, the investigations and prosecutions of some firms were left idle for years. Examples of these are firms that were involved in the grain milling cartel, where some prosecutions were finalised in 2007/08, others in 2010/11 and others in 2019. With information from leniency as well as the conclusion of other prosecutions through settlements, there are limited reasons for the delays in prosecutions.

Third, once a firm involved in collusion knows that its conduct is under investigation, the likelihood of successful prosecution and sanction increases, assuming that the competition authority is competent and effective at investigating and prosecuting cartels. The firm thus has every incentive to minimise the real value of the sanction[64] it faces, subject to the constraints of investigative and prosecutorial efficiency and the effectiveness of the competition authorities.[65] The firm can minimise the sanction in one or a combination of ways. First, the firm could apply for leniency, and if successful, it would face a zero penalty. Second, if not successful in obtaining leniency (e.g. because it was not first to the door of the competition authority), the firm could settle with the competition authority and pay a lower penalty, because of a settlement discount.

Finally, the firm could engage in strategic behaviour that would delay the payment of a penalty for as long as possible, knowing that over time the real value of the sanction will be eroded by inflation. This would effectively equate to a nominal penalty, which would not deter them from ongoing or future anti-competitive behaviour. Similarly, firms could delay the payment of penalties while they deplete the productive assets and revenues on which the penalties are based. This can be achieved through frivolous litigation and forum shopping, where firms approach other courts outside of the normal competition law structures (e.g. the High Court) where they can manage to extract more time. This has the consequence of delaying the penalty

---

[64] Consider a firm's objective penalty minimisation function given by:

$$min\ P_t\ subject\ to\ E\ and\ S$$

where P_t = penalty paid at time t; E = evidence against the firm; S=investigative and prosecutorial efficiency of competition authority (skill).

[65] Efficiency in this case refers to the ability to process investigations and prosecutions in the least amount of time, using the least possible amount of resources. Effectiveness refers to the ability to produce the desired successful prosecution of firms for cartel conduct. It is accepted that in some instances, the duration of investigation and prosecution of cartels is outside the direct control of the competition authority.

such that when it is paid, the revenue base is much smaller, resulting in a lower penalty. Examples include cases in the wire and steel industries, where firms tried to impair reported revenues and even liquidated the legal entities involved in the collusion. Such strategic behaviour can be eliminated or minimised by the Commission taking a proactive approach to the imposition of penalties and reducing delays in the prosecution of firms where collusion is readily found via leniency. Inefficiency on the part of the Commission worsens this situation.

Figure 4.7: Average duration of investigations and prosecutions of firms for collusion
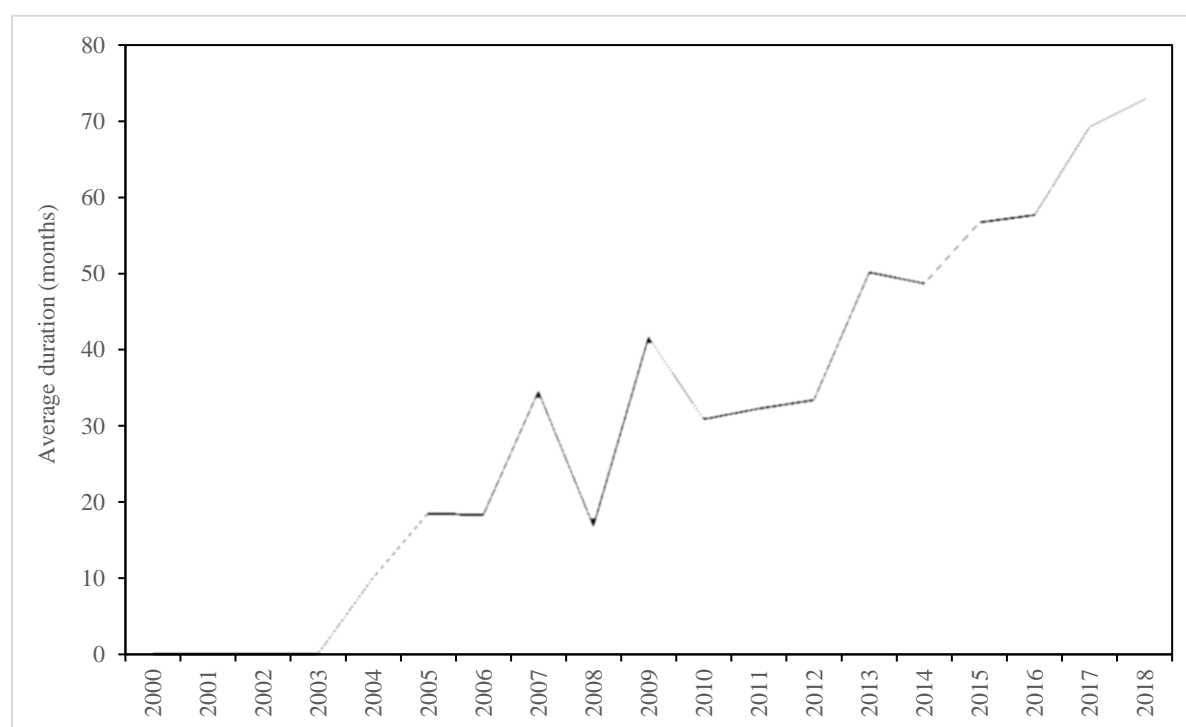
Source: Publicly available Commission and Tribunal records on prosecutions

Table 4.6, below, presents the results of the random-effects estimation of the impact of the number of firms investigated and prosecuted for collusion, as well as the penalties imposed, funding levels for the Commission, and the use of leniency for the duration of investigations and prosecutions of firms for collusion. The random-effects estimation results show that only the funding variable carried the expected sign, but like the other core enforcement variables included in the model, it was not statistically significant. The penalties variable carried a positive sign, and this could be explained by the factors discussed above and the findings of the present-value analysis in the next section. Leniency, which has been key to cartel enforcement, was not associated with reduced durations of cases. The quantitative analysis in Table 4.6 is consistent with the qualitative analysis of the evolution of duration, from

investigation to final prosecution decision, taken together with the evolution of funding, penalties, and the use of leniency.

Table 4.6: Results of random-effects panel regression analysis of the impact of the number of firms, penalties, funding, and leniency on the durations of investigation and prosecution of firms for collusion in South Africa

xtreg: *duration = f (number of firms, penalties imposed, funding of the Commission, use of leniency to detect collusion, trend)*

Bootstrap replications (50)

Random-effects GLS regression				Number of observations = 114		
Group variable: sector				Number of groups = 6		
R-sq: within = 0.7476				Obs per group: min = 19		
between = 0.4342				average = 19.0		
overall = 0.7251				max = 19		
				Wald Chi2(5) = 598.36		
corr (u_i, X) = 0 (assumed)				Prob > Chi2 = 0.0000		
				(Replications based on 6 clusters in sector)		

Duration	Observed coefficient	Bootstrap Std. Err.	Z	P > |z|	Normal-based [95per cent conf. interval]	
Firms	0.0070253	0.1682271	0.04	0.967	-0.3226939	0.3367444
Penalties	0.0282743	0.0194895	1.45	0.147	-0.0099245	0.0664731
Funding	-0.1073202	0.1055823	-1.02	0.309	-0.3142577	0.0996173
Leniency	0.1549930	0.1568096	0.99	0.323	-0.1523482	0.4623341
Trend	0.0827377	0.014163	5.84	0.000***	0.05497880	0.1104966
Constant	1.6727330	1.860689	0.90	0.369	-1.9741520	5.3196170
sigma_u	0.16998612					
sigma_e	0.37068915					
rho	0.17374797 (fraction of variance due to u_i)					

*Statistically significant at: 1 per cent (p<0.01) ***; 5 per cent (p<0.05) **; 10 per cent (p<0.10) ***

The results of the fixed-effects specification in Table 4.7, below, show that none of the variables are statistically significant. Like the random effects specification, the funding variable was negatively related to duration, meaning that more funding can be expected to result in a reduction in duration, as the Commission will have more resources to process investigations and prosecutions more efficiently. However, efficiency will depend on how effectively the Commission deploys its resources.

Table 4.7: Results of fixed-effects panel-regression analysis of the impact of the number of firms, penalties, funding, and leniency on the duration of investigation and prosecution of firms for collusion in South Africa

xtreg: *duration = f (number of firms, penalties imposed, funding of the Commission, use of leniency to detect collusion, trend)*

Bootstrap replications (50)

Fixed-effects GLS regression	Number of observations = 114
Group variable: sector	Number of groups = 6

R-sq: within = 0.7476	Obs per group: min = 19
between = 0.4291	average = 19.0
overall = 0.7248	max = 19

Wald Chi2(5) = 388.08

corr (u_i, Xb) = 0.0068   Prob > Chi2 = 0.0000

(Replications based on 6 clusters in sector)

Duration	Observed coefficient	Bootstrap Std. Err.	Z	P > |z|	Normal-based [95per cent conf. interval]	
Firms	-0.0013743	0.1481734	-0.01	0.993	-0.2917889	0.2890402
Penalties	0.0294205	0.0200596	1.47	0.142	-0.0098955	0.0687366
Funding	-0.1060076	0.1108758	-0.96	0.339	-0.3233202	0.111305
Leniency	0.1504143	0.1514597	0.99	0.321	-0.1464413	0.4472699
Trend	0.0826878	0.0166238	4.97	0.000***	0.0501057	0.1152699
Constant	1.648674	1.951028	0.85	0.398	-2.17527	5.472619
sigma_u	0.16132253					
sigma_e	0.37068915					
rho	0.15923697 (fraction of variance due to u_i)					

*Statistically significant at: 1 per cent (p<0.01) ***; 5 per cent (p<0.05) **; 10 per cent (p<0.10) ***

As in the case of drivers of prosecutions of firms for collusion above, the Blundell and Bond (1998) dynamic panel data model was used to address potential endogeneity issues between the duration of investigations and leniency, as well as between the duration of investigations and number of firms prosecuted. Table 4.10 in the Appendix shows the results of this systems GMM estimation. The results show that leniency has been associated with a smaller positive relationship with the duration of investigations, which is not statistically significant. The size of the coefficient under the systems GMM specification is smaller than that under the standard random-effects and fixed-effects specifications despite having the same sign and not being statistically significant across all specifications. This confirms that leniency has not contributed

133

to a reduction in the duration of investigations. Funding on the other hand, appears to contribute to a reduction in the duration of investigations. Under the systems GMM specification, funding has a larger statistically significant negative relationship with the duration of investigations compared to the random-effects and fixed effects specification results which are not statistically significant. Caution should be exercised when interpreting the coefficient for funding regardless of its statistical significance given the limitations associated with the variable as explained above.

The results of the study show that while leniency and settlements have played a major role in cartel detection and prosecution in South Africa, they do not appear to have raised efficiency[66] levels in investigation, prosecution, and penalisation of firms for collusion. It is also important to note that the Commission's manpower has increased significantly compared with the earlier years, with a standalone unit for investigating cartels. It appears the Commission's resources are tied up in cartel investigations despite the availability of leniency and settlement procedures as enforcement tools available to the Commission. The observations on the duration of cartel cases are akin to those in the study by Ascione and Motta (2010). In Europe, for cases involving leniency, Ascione and Motta (2010) reported duration of investigations ranging from 13 months to 58 months. The authors conclude that the resources of the EC are tied up in cartel cases, even when leniency yields key evidence with cases rarely lasting less than three years. For those cases that did not involve leniency, the authors reported the duration of investigations as ranging from 14 months to 62 months. This suggests that one of the central benefits of having leniency and settlement procedures (time and resource saving) have not been realised, despite the two instruments being central to cartel enforcement in South Africa.

The timely prosecution of firms for collusion is as important as the prompt detection of collusion because delays in prosecution undermine the effectiveness of penalties in enhancing

---

[66] In this study, efficiency is measured as the duration of cases from detection to prosecution and penalisation, all else given. There may well be legitimate circumstances/factors that drive up duration, e.g. a dispute on a legal principle, which entails a case being litigated in various courts. It is accepted that these are legitimate elements of law enforcement since competition authorities are themselves subject to the law. In such circumstances, the existence of an infringement and the resulting penalty are determined by the courts and not by settlement between the competition authority and the firm. There are instances where the competition authority and the firm agree to suspend the determination of certain elements of prosecution (e.g. sanction) pending the determination of legal questions by the courts. Where this has occurred, the study examined the efficacy of such agreements from a deterrence perspective. For example, there appears to be limited rational reason for the Commission to enter an agreement to suspend the determination of a penalty when the dispute is about whether the conduct amounts to that of a cartel. The Commission could prosecute both with its motivation for a penalty, and the court could simultaneously decide on both issues, thereby reducing delays in the imposition of penalties.

deterrence. Firms may even delay cases and then settle them later to enjoy a discount for settlement. This has implications for the efficacy of a deterrence-based framework. It is worth pointing out that in South Africa, even the longest running cases are mostly resolved through settlement, meaning that firms effectively get rewarded for delays in finalising cases, instead of being penalised. At this point there is no strong public evidence from final settlement agreements that the Commission strongly advocates for increased penalties for firms that delay the settlement of cases.

This may also have the perverse effect of encouraging stable cartels to continue colluding, knowing that if detected, sanction will be delayed and could be reduced by entering into a settlement agreement with the Commission at a later stage. Firms that would also ordinarily settle cases early will also learn that early settlement results in higher real penalties compared with later settlement, which will have the unintended effect of disincentivising firms from settling cases early. These aspects of the cartel enforcement record often get overlooked in the pomp and funfair of the nominal penalties collected by the Commission and the media statements that reveal the identity, nature and size of the firms penalised for collusion, as well as the public anger at the depth and scale of collusion.

If cartel conduct leads to more permanent or enduring damage to markets long after the infringement has ended, delays in penalties taken together with the enduring damage may fail to dissuade firms from violating the law in the first place, even if the size of the sanction may appear high in nominal terms. The profit from violating the law includes the benefits that firms obtain long after the contravention has ended if it takes time before the market can be effectively competitive again. Both the medium-to-long-term impact of the effectiveness of competition in markets and the likelihood of timely detection and punishment are important elements of whether deterrence is achieved.

The value of delaying finalisation and payment of penalties is therefore important to understand through the lens of firms involved in collusion, and the implications that this has for deterrence. To do this, the present-value technique was applied in the study, as discussed below.

4.5.3 Impact of duration of investigation, prosecution, and penalisation on the value of penalties paid by firms for collusion and on deterrence

In applying the present-value technique, the present values of penalties imposed on firms for collusion during the period were calculated. Table 4.8, below, presents selected summary

statistics relating to discounts for delays in investigations and prosecutions of firms for collusion, and the associated implications for the present value of penalties. The average duration of cases for the period under study was 5.5 years, and the average time-related discount on penalties was 34 per cent. In other words, the value of penalties paid by firms for involvement in collusion on average was 34 per cent lower in present value terms when discounted using the average discount rate on five- to 10-year government bonds. The shortest case duration (0.8 years) was associated with the lowest discount on the present value of penalties (seven per cent) and the longest case duration, lasting 11.7 years, was associated with the largest present value discount, of 61 per cent.

Another way to show the impact of delays on the penalty paid is to consider how the delay changes the penalty as a percentage of revenue, as well as how the delays impact the relative penalties paid by two or more firms involved in the same cartel. In the study, the example of Pioneer Foods – which opted for litigation and delayed payment of the penalty by about three years – is used to show the impact of delays on the penalty paid as a percentage of turnover. When Pioneer Foods paid the penalty, the penalty was set at R195.7 million which amounted to 9.88 per cent of its bread division's turnover for 2006. The present value of this penalty was R151.0 million which amounted to 7.62 per cent of the bread making division's turnover for 2006. Tiger Brands paid, within about a year, a penalty that amounted to 5.7 per cent of its turnover for bread. The discounted value of the penalty amounted to 5.32 per cent of its turnover for bread, reflecting a 0.38 per cent time-related discount on the penalty. This shows that by delaying payment of the penalty by about three years, Pioneer Foods managed to shave off a time-related 2.26 per cent points of the penalty as a percentage of turnover, while a shorter delay by Tiger Brands, for example, would have resulted in a smaller time-related discount of 0.38 per cent. Both Tiger Brands and Pioneer Foods were involved in the bread cartel but resolved their cases at different times using different mechanisms. Tiger Brands settled its case and Pioneer Foods opted to litigate in a case that took about three years. The present value of the penalty paid by Tiger Brands within a year resulted in a present value discount of seven per cent while the three-year delay by Pioneer Foods resulted in a present value discount of 23 per cent on the penalty paid. Again, this shows that delays diminish the real quantum and effectiveness of the penalty.

Table 4.8: Summary statistics on delays in finalisation of cases and the impact it has on the present value of penalties paid

Variable	Value
Shortest case duration for penalised firms	0.8 years
Average case duration for penalised firms	5.5 years
Longest case duration for penalised firms	11.7 years
Average penalty paid	R45.6 million
Average present value	R32.6 million
Average penalty discount	R12.9 million
Lowest percentage penalty discount for shortest case duration	7 per cent
Average percentage penalty discount across prosecutions	34 per cent
Highest percentage penalty discount for longest case duration	61 per cent

Source: Own calculations based on cartel prosecution data and SARB interest rates

Table 4.11 in the Appendix presents the results of the analysis. As a general observation, and consistent with observations in the finance field, the present value of penalties paid in the future declines with time. This means that firms that settle cases and pay penalties earlier effectively pay higher penalties compared with those that delay penalties. Stated differently, firms that delay paying penalties receive a greater time-related discount on the present value of the penalties they eventually pay. Firms that can better negotiate lower penalties with the Commission can also have lower discounted penalties if they pay the penalty earlier. The results in Table 4.11 (reported in the Appendix) show few instances[67] where the discounts for penalties paid after shorter periods are greater than those paid after longer periods. This provides reassurance that the observations on the value of delays to firms holds in general.

Given the observations from the cartel enforcement record and by applying time-value-of-money techniques, and interfacing this with the notion of deterrence, which is central to competition law enforcement, the key question is why does the Commission rely on settlements and leniency in the manner that it does? The question is highly relevant given that settlement and leniency do not appear to have contributed significantly to procedural economy. There could be several reasons for the Commission relying on settlements despite the foregoing. First, settlements are easier than full litigation, especially where the outcome for the Commission is uncertain on the quantum of the penalties. Where firms are prepared to settle at any point in time, the Commission would rather negotiate as high a penalty as possible, rather than insisting

---

[67] This applied to only five out of 75 prosecutions involving penalties on firms.

on litigating. In the end, the outcome depends on who – between the Commission or the firm – is a more skilled and better negotiator.

The second reason for the Commission's use of settlement and leniency under these circumstances may be concerned about prosecutions being prolonged further, as shown by the record, forcing it to settle, rather than to continue litigating. In other words, litigation fatigue could result in the Commission opting to close the file through settlement, rather than to continue the prosecution through litigation. While this may be appealing, it has the potential to undermine deterrence if it results in under-deterring penalties and may reduce procedural efficiencies when cases are not resolved quickly. For example, there have been cases where a firm accepts that it violated the law but fails to reach agreement with the Commission on the penalty to be paid, leading to prolonged litigation that is ultimately resolved by settlement with a penalty that is lower than what the Commission initially demanded.

The third reason for the Commission's approach may be because several cases have either been overturned or have reduced penalties for firms when they go on appeal. As a result, the Commission might see it as being better to settle a case and avoid the risk of either having cases overturned or penalties being reduced on appeal. In other words, the Commission may well have adopted a more risk-averse approach. There are many legitimate reasons for cases to be overturned or penalties to be reduced. The suggestion here is not that the appeals have no merit, but that the impact of appeal-related reductions in penalties and the overturning of decisions might negatively impact the Commission's appetite to litigate cases, especially where this can be avoided.

Finally, settlement brings results, which the Commission can show to both policymakers and the public, and because no work has until now been undertaken to ascertain and show whether deterrence is undermined, there is room for rent-seeking behaviour by the Commission, by chasing after settlement in a way that effectively disregards the medium- to long-term impact on deterrence.

4.6   Conclusion

In a deterrence-based framework of cartel enforcement, competition regulators are concerned with two elements: first, their ability to detect cartels, p; and second, their ability to penalise them. Given that it is impossible to have absolute deterrence, where $p = 1$, competition authorities look to enhance levels of p and F through policy initiatives such as the use of

leniency policies, settlement procedures, increasing their capacity to investigate and prosecute cartels (both head count and skills set), and advocacy, which raises awareness of the cartel problem. These tools complement one another, but also involve trade-offs that need to be carefully considered if deterrence is to be achieved.

In this study, the relationship between penalties, leniency and the Commission's funding, on the one hand, and the number of firms prosecuted for involvement in collusion, on the other was examined. The focus was on the 20 years since the introduction of the current competition law regime. The results show that under an OLS-panel specification, the level of penalties, leniency and the funding of the Commission have a positive, statistically significant effect on the number of firms prosecuted. These results suggest that providing the Commission with more funding, coupled with the Commission implementing policies that enhance cartel detection and punishment will reduce the incidence of collusion. The caution here is, of course, that the cartel universe is unknown, and this might have implications for the generality of the results.

Leniency and settlement procedures are expected to enhance cartel enforcement efficiency by placing cartel members in asymmetric positions, leading to whistleblowing and thereby aiding cartel detection. These procedures therefore reduce the amount of time and resources spent on investigating and prosecuting cartels. The analysis of the duration of cartel investigations and the prosecution of firms for collusion showed that over time, the duration of investigation and prosecutions have increased significantly compared with the period prior to the introduction of the leniency policy, in 2004, and its revisions, in 2008. This is contrary to expectations, given that most cases in the period of study were resolved by settlement, and a significant proportion involved leniency. Several reasons can be offered for this. First, the number of cartel cases has increased significantly, and cartel cases have become more complex. For example, cartel investigations often involve industry-wide collusion, requiring complex arrangements to resolve (e.g. the construction-sector collusion). Relatedly, leniency applicants may not be yielding water-tight hard evidence that enables the Commission to readily prosecute other cartel firms, implying that the Commission spends more time and resources searching for more evidence.

The second reason for the increased duration of prosecutions subsequent to the introduction of a leniency policy is that the Commission has not been effective in resolving cartel cases timeously, even where leniency has been granted. For instance, some of the cases resolved by

the Commission in 2018 are from the 2006/07 to 2007/08 period. In these cases, the investigations and prosecutions of some firms remained idle for years without any publicly observable policy action (e.g. litigation in courts) or stated reasons in public documents released by the Commission. For example, the grain milling cartel case was resolved over a long period, with some prosecutions finalised in 2007/08, others in 2010/11 and others only in 2019. With information from leniency, as well as the conclusion of other prosecutions through settlements, there seem to be limited reasons for the delays in prosecutions.

The third reason for the increased duration of prosecutions is the strategic behaviour by firms aimed at paying time-discounted penalties. Firms know that the real value of the monetary sanction is eroded over time by factors such as inflation, and where firms set aside a smaller amount towards the penalty, a significant proportion of the penalty ultimately paid can be generated from interest earned from financial markets. This effectively represents nominal penalties, especially because the Commission does not adjust penalties for the time that elapses between cartel detection and settlement. If penalties are not adjusted to reflect these delays, it undermines the deterrence effect of penalties in a manner that may even offset the positive impact of leniency and settlements. The negative impact on deterrence is worsened by the ineffectiveness of the Commission at resolving cases timeously.

Firms delay by waiting for investigations to start and get to an advanced stage before either applying for leniency or engaging in settlement negotiations, and still receive a settlement discount.[68] Similarly, as observed in some cases, firms could delay paying penalties while they deplete the productive assets and revenues on which the penalties are based. This can be achieved through frivolous litigation and forum shopping, where firms approach other courts outside of the normal competition law structures (e.g. the High Court) where they can extract more time. This has the consequence of delaying the penalty, such that when it is paid, the revenue base is much smaller, resulting in a lower penalty. Examples of this include cases in the wire and steel industries, where firms tried to impair reported revenues and even liquidated the legal entities involved in collusion. Such strategic behaviour can be eliminated or minimised by the Commission taking a proactive approach to imposing penalties, and by reducing delays in prosecutions where collusion is detected via leniency.

---

[68] Firms can apply for leniency even after the Commission has started investigations.

While leniency and settlements have played a major role in detecting cartels and closing investigations, they have not increased efficiencies in prosecuting and penalising firms. The results of this study show that the duration of investigations and sanctions has been increasing over time, and not decreasing as would have been expected in theory. The timely prosecution of firms for collusion is as important as the timely detection of collusion because delays in prosecution undermine the effect of penalties in terms of deterrence. This has implications for the efficacy of a deterrence-based framework, especially where cartel prosecutions are mostly resolved through settlement. This may have the unintended effect of rewarding firms for delays in finalising cases, instead of penalising them. It also has the perverse effect of encouraging stable cartels to continue colluding, knowing that if detected, sanction is delayed and can be reduced by late settlement. Also, firms that would ordinarily have settled early could learn that early settlement results in higher real penalties, compared with later settlement, which would have the effect of disincentivising firms from settling early. This challenge for South African cartel enforcement is best revealed by a critical review of the enforcement record, as presented in this chapter. The challenge of delays undermining deterrence is often overshadowed by both the attention paid by the popular media to the identity, nature and size of the firms penalised for collusion and the accompanying public discontent over the depth and scale of collusion.

There are several policy implications that flow from this study. First, some cartel prosecutions involve settling cases where firms have violated competition laws in other markets (e.g. collusion in milling and bread markets) and where leniency applications have been received. Examples of these include collusion in the bread, milling, steel and construction markets. In such instances, the Commission should consider ensuring that penalties imposed on firms for bundled settlements take into consideration the scope of the collusion as well as delays in finalising the cases. This requires a proactive approach to penalties, settlements, and leniency.

To ensure that deterrence is effective from a settlements and penalties perspective, competition authorities should look to expedite the finalisation of investigations, prosecutions, and imposition of penalties. To the extent that delays are outside the control of the Commission, the penalties imposed should be set at levels higher than those for firms settle early. It is important to note that most of the cartel cases handled by the Commission are settled, and not contested litigation cases. Competition authorities can also consider adding interest to the penalty amount for the period between the year for which the penalty is calculated and when it is paid. This enhances deterrence of penalties. The results of this study, particularly those

derived from applying time-value of money techniques, suggest that there is a need for the competition authorities (comprising the Commission, the Tribunal and the CAC) to review and reconsider their approach to penalties, and to shift their attention to deterrence.

For the Commission, this means that when deciding and negotiating settlement agreements, it should particularly consider insisting on higher effective penalties for delayed settlements, especially where there are no legitimate developments in cases that would have occasioned the delays. Firms that delay settlements ought to face penalties that are higher, in real terms, than those that settle cases earlier. This can be achieved by implementing a clear policy where settlement discounts are progressively reduced as time passes. This should apply without firms having to settle cases such that at some point they do not enjoy the benefits of the settlement discounts. This may especially be the case if firms do not settle within the first year of the cartel being detected. This policy approach will ensure that firms have the incentive to settle early rather than delaying. It will also help to ensure that the expected benefits of having leniency and settlement procedures are preserved, because the longer a case takes and the more effort, resources and time that the Commission puts into a case, the less likely that settlement will generate the anticipated benefits. At this point, there is no consistent demonstrable evidence in all the cases that have been concluded that delays in settlements have been penalised through reduced settlement discounts. This may well be because most cases are resolved through settlements that are outcomes of bargaining between the Commission and firms. The bargaining often starts with the firm proposing a penalty to the Commission, the lowest possible penalty that it would be willing to pay. This is tantamount to markedly rewarding firms that settle late and consume the Commission's time and resources, and this undermines deterrence.

On its part, the Commission ought to ensure that it cuts the prevalence of cases that lie fallow after other cartel members have settled. Examples of such cases include those in the bread and milling markets, which have been idle for years subsequent to the settlement of cases by other cartel members, without any contestation by the firms that had not yet settled. These are apparent inefficiencies on the part of the Commission. The Tribunal also should consider tightening its role and approach to settlement agreements, by taking a more proactive approach towards holding both the Commission and the firms being penalised accountable for the penalties proposed in settlement agreements. This role is of course subject to the limitation that the Tribunal is an adjudicative body with powers that are determined by statute.

The CAC, on the other hand, when approached with appeals relating to the quantum of the penalty, should consider the extent of time delays in determining the penalties to be imposed, and should assess whether there are legitimate causes for delays, rather than opportunistic behaviour. This may be a controversial issue given that firms have a legitimate right to protect and defend their interests, and the exercising of such rights should not be viewed as unreasonable and punishable with higher penalties. Exercising such rights should be seen in the context of the need to achieve deterrence, and only legitimate defences of the rights of firms should be allowed to delay the finalisation of cases. As mentioned previously, there have been cases where firms delay cases by engaging in frivolous litigation in different court systems, and this has continued for several years. Such behaviour should be discouraged through higher penalties.

The legal system should also proactively encourage firms to come forward and settle cases with the Commission. This could entail actively implementing a system of penalties and discounts for settlement. The Commission's guidelines for determining penalties envisages this, but up to this point, the evidence has not been clear that the principles set forth are being applied in all cases. These are not easy issues to address, but it is possible for the Tribunal and the CAC to make rational judgment calls on the legitimacy of issues raised in cases. Courts routinely weigh up evidence and assess the conduct of firms to determine the validity of claims and the legitimacy of behaviour, and such findings must be clearly reasoned and articulated in judgments to signal a clear change in approach. Reviews and appeals should not be allowed to be used as opportunistic instruments for lowering penalties and to unduly delay the finalisation of cases in order to dilute the impact of penalties. Such behaviour ought to be discouraged so that the incentives to firms are consistent with the aim of competition law, which is to deter firms from colluding.

Areas for further study include developing tools to better understand the incentives of firms to settle, and how firms delay paying penalties in this context. This is a dynamic process, which is not captured by the static considerations that characterise much of the theory and policy debates. There has been some work on understanding the incentives for firms and competition authorities to resolve cases through settlement procedures, and a substantial part of this research has been limited to the desire to save costs. Not much attention has been paid to the time dimension, which influences the willingness and timing of settlements between the competition authorities and firms. This has implications for what would constitute an optimal

cartel enforcement policy in South Africa. The author argues that an optimal cartel enforcement policy would be one that considers discount rates for penalties and uses sanctions in a manner that dynamically incentivises early settlement or finalisation of cases by progressively and more aggressively penalising delays in settling cases. This is especially the case where the cartel enforcement record is characterised by significant delays, often lasting several years, in settling and finalising cases.

APPENDIX

Table 4.9: Results of GMM panel regression analysis of the impact of penalties, funding, and leniency on the number of firms prosecuted for collusion in South Africa

xtdpdsys: *number of firms prosecuted = f (penalties imposed, funding of the Commission, use of leniency to detect collusion), robust*

System dynamic panel-data estimation				Number of observations = 108	
Group variable: sector				Number of groups = 6	
Time variable: year					
				Obs per group: min = 18	
				average = 18	
				max = 18	
Number of instruments = 85				Wald Chi2(4) = 742.88	
				Prob > Chi2 = 0.0000	

Firms	Observed coefficient	Robust Std. Err.	Z	P > |z|	[95per cent conf. interval]	
Firms (lag 1)	-0.0118051	0.0738324	-0.16	0.873	-0.156514	0.1329037
Penalties	0.0762079	0.0077113	9.88***	0.000	0.0610940	0.0913219
Funding	0.1201338	0.1081013	1.11	0.266	-0.091741	0.3320085
Leniency	0.3612205	0.0891094	4.05***	0.000	0.1865693	0.5358717
Constant	-2.203364	1.979087	-1.11	0.266	-6.082303	1.6755760

*Statistically significant at: 1 per cent (p<0.01) ***; 5 per cent (p<0.05) **; 10 per cent (p<0.10) ***

Table 4.10: Results of GMM panel regression analysis of the impact of the number of firms, penalties, funding, and leniency on the durations of investigation and prosecution of firms for collusion in South Africa

xtdpdsys: *duration = f (number of firms, penalties imposed, funding of the Commission, use of leniency to detect collusion, trend), robust*

System dynamic panel-data estimation				Number of observations = 108	
Group variable: sector				Number of groups = 6	
Time variable: year					
				Obs per group: min = 18	
				average = 18	
				max = 18	
Number of instruments = 101				Wald Chi2(5) = 21.41	
				Prob > Chi2 = 0.0007	

Duration	Observed coefficient	Robust Std. Err.	Z	P > |z|	[95per cent conf. interval]	
Duration (lag 1)	0.2011604	0.0759655	2.65	0.008	0.0522708	0.3500500
Firms	-0.0029772	0.1253959	-0.02	0.981	-0.2487434	0.2427997
Penalties	0.0299354	0.0148344	2.02	0.044**	0.0008604	0.0590103
Funding	-0.4636224	0.1840989	-2.52	0.012**	-0.8244497	-0.1027951
Leniency	0.0532201	0.1236850	0.43	0.667	-0.1891980	0.2956381
Trend	0.1409171	0.0344340	4.09	0.000***	0.07342770	0.2084065
Constant	7.7786310	3.2322320	2.41	0.016**	1.44357400	14.11369

*Statistically significant at: 1 per cent (p<0.01) ***; 5 per cent (p<0.05) **; 10 per cent (p<0.10) **

Table 4.11: Present value of penalties for collusion

Respondent	Case duration (years)	Penalty	Average discount rate	PV of penalty	Penalty discount caused by delay	Penalty discount caused by delay
Tiger Brands	0.8	R98,784,870	8.2%	R92,241,224	R6,543,646	7 %
Foodcorp	2.0	R45,406,360	8.4%	R38,350,457	R7,055,903	16 %
SAA, SA Airlink, SA Express	2.1	R20,000,000	8.3 %	R16,833,545	R3,166,455	16 %
Trident Steel	2.8	R8,563,836	7.6 %	R6,899,159	R1,664,676	19 %
Keystone Milling Co.	3.0	R6,730,349	8.4 %	R5,228,994	R1,501,355	22 %
Pioneer Foods	3.1	R195,718,614	8.4 %	R150,996,801	R44,721,813	23 %
Amalgamated Metals Recycling CC	3.2	R3,264,945	8.3 %	R2,510,642	R754,302	23 %
Abeddac Metals CC	3.2	R4,965,794	8.3 %	R3,818,543	R1,147,250	23 %
Universal Recycling Company	3.2	R18,061,597	8.3 %	R13,888,815	R4,172,782	23 %
McCoys Glass Wholesalers CC	3.3	R2,487,451	7.6 %	R1,943,173	R544,277	22 %
Afrisam South Africa	3.4	R124,878,870	8.1 %	R94,575,771	R30,303,099	24 %
Glass South Africa	3.4	R4,395,023	7.6 %	R3,390,147	R1,004,876	23 %
National Glass Distributors	3.4	R414,615	7.6 %	R319,818	R94,797	23 %
Pioneer Food	3.7	R660,000,000	8.3 %	R487,688,429	R172,311,571	26 %
Lafarge Industries South Africa	3.8	R148,724,400	8.1 %	R109,621,733	R39,102,667	26 %
Power Metals	3.8	R12,773,588	8.3 %	R9,309,763	R3,463,825	27 %
Northern Hardware and Glass	3.9	R214,531	7.6 %	R159,312	R55,219	26 %
Lufthansa	3.9	R8,500,000	9.1 %	R5,953,514	R2,546,486	30 %
SAA	3.9	R20,000,000	9.1 %	R14,008,269	R5,991,731	30 %
Carolina Rollermeulle	4.0	R4,417,546	8.2 %	R3,186,725	R1,230,821	28 %
Stefanutti Stocks Holdings	4.0	R55,864,536	7.6 %	R41,223,501	R14,641,035	26 %
Hochtief Construction AG	4.2	R1,907,793	7.6 %	R1,390,082	R517,711	27 %
National Scrap Metals Cape Town	4.3	R17,730,974	8.3 %	R12,486,086	R5,244,888	30 %
Singapore Airlines	4.3	R25,106,692	7.9 %	R17,961,719	R7,144,973	28 %
Basil Read Holdings	4.4	R94,936,248	7.8 %	R67,232,737	R27,703,511	29 %
G Liviero & Son Building	4.4	R2,011,078	7.8 %	R1,424,222	R586,856	29 %
Haw & Inglis Civil Engineering	4.4	R45,314,041	7.8 %	R32,090,872	R13,223,169	29 %
Hochtief Construction AG	4.4	R1,315,719	7.8 %	R931,777	R383,942	29 %
Norvo Construction	4.4	R714,897	7.8 %	R506,282	R208,615	29 %

Respondent	Case duration (years)	Penalty	Average discount rate	PV of penalty	Penalty discount caused by delay	Penalty discount caused by delay
Raubex	4.4	R58,826,626	7.8 %	R41,660,326	R17,166,300	29 %
Rumdel Construction Cape	4.4	R17,127,465	7.8 %	R12,129,470	R4,997,995	29 %
Stefanutti Stocks Holdings	4.4	R306,892,664	7.8 %	R217,337,783	R89,554,881	29 %
WBHO Construction	4.4	R311,288,311	7.8 %	R220,450,728	R90,837,583	29 %
Murray & Roberts	4.4	R309,046,455	7.8 %	R218,863,072	R90,183,383	29 %
Esorfranki	4.4	R155,850	7.8 %	R110,371	R45,479	29 %
Vlaming	4.4	R3,421,662	7.8 %	R2,423,181	R998,481	29 %
Tubular Technical Construction	4.4	R2,634,667	7.8 %	R1,865,840	R768,827	29 %
Aveng Africa	4.4	R306,576,143	7.8 %	R217,113,626	R89,462,517	29 %
Guiricich Bros Construction	4.4	R3,552,568	7.8 %	R2,515,887	R1,036,681	29 %
WBHO Construction	4.7	R10,244,136	7.6 %	R7,186,837	R3,057,299	30 %
Harding Allison Close Corporation	4.8	R78,822	7.6 %	R54,949	R23,873	30 %
B and E International	4.8	R8,158,447	7.6 %	R5,651,610	R2,506,837	31 %
Cycad Pipelines	4.9	R3,394,151	7.6 %	R2,336,399	R1,057,752	31%
N17 Toll Operators	4.9	R424,121	7.6 %	R291,948	R132,173	31 %
Civcon Construction	5.2	R798,386	7.6 %	R539,240	R259,146	32 %
Isipani Construction	5.4	R10,280,295	7.7 %	R6,765,458	R3,514,837	34 %
Giuricich Coastal Projects	5.5	R149,429	7.6 %	R98,403	R51,026	34 %
Foodcorp	5.7	R88,500,000	8.0 %	R56,311,562	R32,188,438	36 %
Pele Kaofela	5.8	R437,624	7.7 %	R281,196	R156,427	36 %
Enviroserve	5.8	R10,209,519	7.9 %	R6,483,301	R3,726,218	36 %
South African Airways	6.2	R18,799,292	8.0 %	R11,494,000	R7,305,292	39 %
Murray and Roberts Ltd	6.3	R64,141,799	7.7 %	R39,406,033	R24,735,766	39 %
Dura Soletanche-Bachy	6.4	R988,589	7.7 %	R603,467	R385,122	39 %
Air France – KLM	6.5	R18,634,640	8.0 %	R11,094,327	R7,540,314	40 %
British Airways Plc	6.5	R12,204,730	8.0 %	R7,266,213	R4,938,517	40%
British airways Plc	6.5	R21,765,297	7.9 %	R13,063,444	R8,701,853	40 %
Geomechanics CC	7.3	R1,650,503	7.8 %	R931,609	R718,895	44 %
Delatoy Investments	7.4	R4,136,122	7.9 %	R2,298,820	R1,837,302	44 %
Guiricich	7.7	R900,000	7.8 %	R494,179	R405,821	45 %
Martinair Cargo	7.9	R5,758,250	7.9 %	R3,077,136	R2,681,114	47 %
Cargolux International SA	8.7	R11,000,000	7.9 %	R5,557,236	R5,442,764	49 %

Respondent	Case duration (years)	Penalty	Average discount rate	PV of penalty	Penalty discount caused by delay	Penalty discount caused by delay
Rodio Geotechnics	8.8	R885,963	7.9 %	R441,158	R444,805	50 %
Edilcon Construction	8.9	R10,510,680	7.9 %	R5,199,398	R5,311,282	51 %
SA Metal Group	9.0	R22,430,000	7.9 %	R10,998,844	R11,431,156	51 %
Ton Scrap	9.7	R3,500,000	8.0 %	R1,619,069	R1,880,931	54 %
Cape Gate	9.8	R40,000,000	8.0 %	R18,243,695	R21,756,305	54 %
Blinkwater Mills	10.0	R10,112,504	8.0 %	R4,555,238	R5,557,267	55 %
Ben Jacobs Iron and Steel	10.1	R2,995,923	8.0 %	R1,340,592	R1,655,330	55 %
Bothaville Milling	10.3	R4,211,386	8.0 %	R1,859,451	R2,351,935	56 %
Brenner Mills	10.3	R12,000,872	8.0 %	R5,263,637	R6,737,235	56 %
Pride Milling	10.5	R10,624,960	8.0 %	R4,598,619	R6,026,341	57 %
Godrich Milling	10.6	R4,354,467	8.0 %	R1,872,184	R2,482,283	57 %
Progress Milling	11.3	R2,120,000	8.0 %	R859,922	R1,260,078	59 %
TWK Milling	11.7	R1,845,864	8.0 %	R724,117	R1,121,747	61 %

Source: Own calculations based on cartel prosecution data and SARB interest rates

Table 4.12: Results of random and fixed-effects panel-regression analysis of the impact of penalties, funding, caseload, and leniency on the number of firms prosecuted for collusion in South Africa

xtreg: *number of firms prosecuted = f (penalties imposed, funding of the Commission, use of leniency to detect collusion)*

Bootstrap replications (50)

Firms	Random effects incl. caseload and funding	Fixed effects incl. caseload and funding	Random effects excl. funding	Fixed effects excl. funding
Penalties	0.07448	0.07272***	0.07571***	0.07407***
	(8.62)	(7.57)	(8.62)	(8.54)
Funding	0.07724	0.07768		
	(1.14)	(1.05)		
Caseload	-0.02685	-0.02291	0.11241	0.11306
	(-0.19)	(-0.16)	(1.22)	(1.31)
Leniency	0.36533***	0.38524***	0.37300***	0.39553***
	(4.63)	(3.89)	(5.36)	(4.05)
Constant	-1.23384*	-1.26313	-0.69158	-0.69402
	(-1.67)	(-1.52)	(-1.22)	(-1.31)

*Statistically significant at: 1 per cent (p<0.01) ***, 5 per cent (p<0.05) **, 10 per cent (p<0.10) ***

Chapter 5

Conclusion

Extensive collusion has been detected in South Africa over a period of more than 20 years, in many instances with firms participating in collusion over several years. Some firms participated in cartels in multiple markets or belonged to a group of firms in which another firm was also prosecuted for collusion. Some of the key cartels in South Africa cartelised multiple products within the same market. Overcharge estimates from studies of various South African cartels show that the cartels have been able to raise prices above competitive levels for a sustained period. These justify active cartel enforcement by competition authorities. The Commission has received widespread recognition for its successes in enforcement activities against cartels, mostly measured in terms of number of cartels detected, penalties collected and evidence that it has stopped harmful conduct based on estimated overcharges.

Despite all this, no study has examined the cartel record with a view to understanding the characteristics of collusion and the institutional features of selected key cartels in South Africa. Similarly, there are limited empirical studies examining the pricing dynamics of multi-product firms participating in multi-product cartels that have operated over different collusive phases intermingled with non-collusive phases. The cartel enforcement record has also not been examined comprehensively across its different components to critically evaluate its efficacy from a deterrence perspective. With the above background in mind, this dissertation set out to answer the following three broad questions:

(i)   Research question 1: What are the characteristics of detected South African cartels, and what are the institutional features of selected key cartels?

(ii) Research question 2: What is the pricing behaviour of multi-product firms involved in collusion and the extent of harm that they impose on consumers?

(iii) Research question 3: What have been the key drivers of cartel enforcement in South Africa, and how effective has cartel enforcement been when examined through the lens of efficiency and deterrence?

These questions are answered in Chapters 2, 3 and 4, respectively, of this dissertation. The main findings of the dissertation are summarised in the next section.

5.1   Summary of main findings

5.1.1 Chapter 2: Characteristics of South African cartels focusing on institutional features of selected key cartels

Selected key cartels detected in South Africa have relied on many of the same institutional structures that the theoretical and empirical literature show as aiding collusion. Detected cartels in South Africa, as predicted in the theoretical and empirical literature, have had to navigate challenges ranging from difficulties in aligning disparate interests and cheating to the threat of entry or expansion by firms outside the cartels. Evidence from the cartel enforcement record shows that these difficulties do not make cartels inherently unstable and that South African cartels often endure for several years.

Four institutional features of major South African cartels were studied. The first, communication, which is central to collusion, seems to have involved a mix of different forms, comprising meetings, bilateral telephonic and email communication and, in some instances, extensive information-sharing arrangements. In some cartels, cheating is detected through customer relationships and price-matching clauses. This suggests that the basic sales infrastructure at the level of individual firms could be used to support collusion. This infrastructure complements imperfect cartel-level information exchanges, enhancing the chances of successful collusion. Cartels that shared extensive information, including aggregated, less-frequent production cost information developed more entrenched knowledge of their rivals' businesses. If production processes are not dynamic (involving continuous change), rivals will have a greater chance of monitoring fellow cartel members' pricing behaviour, even with imperfect information on discounting. The firms that faced more complex challenges appear to have enhanced the collusive mechanism, e.g. by agreeing on list prices and discounts for different classes of customers.

The second is the vertical nature of collusion in some of the cartels. This feature suggests that such arrangements existed to stabilise and make collusion more effective, by enhancing the ability to align interests, keep incentives intact, deal with external threats more effectively in a co-ordinated way, and enhance the exercising of cartel market power along the value chain. Ordinarily, a cartel at one level of the value chain is harmful to the interests of the other level, by both reducing demand and raising the costs of production. As such, firms at one level are expected to be averse to having a cartel either at the downstream or upstream level. Upstream

cartel members were suppliers of key inputs to downstream competitors, including those outside the collusive ring. For most of these downstream rivals, the next best alternative source of inputs was imports. In this area, the input markets were in some instances protected using import-parity pricing and trade tariffs, instituted at the request, or with the support of cartel members or the trade associations representing the cartel members. Trade policy tools were costless mechanisms for limiting the threats posed by imports.

The third feature relates to how some South African cartels shared economic rents from collusion. The results of the study show that South African cartels used a wide variety of mechanisms to share the cartels' rents. These included market or customer allocation, joint ventures, sub-contracting, and some form of payment schemes. Interaction across multiple markets and over time helps cartels to simultaneously share rents and institute overarching structures that subsist over time. Collusion in the construction industry shows that the quid pro quo nature of collusion is best understood when one examines cartel behaviour at a macro level, with individual contracts/projects reflecting how rents were distributed across firms.

The fourth feature relates to how cartels dealt with threats of entry and expansion. There was some evidence suggesting that some cartels used trade policy tools strategically to limit the destabilising threat of entry and expansion by firms outside the collusive arrangement. Entry and expansion using imported inputs undermines the collusive price of upstream inputs and erodes the downstream profitability of collusion.

5.1.2 Chapter 3: Pricing dynamics of a multi-product cartel

Multi-product cartels priced their products to maximise profits, given demand and supply conditions. As a result, one could have expected overcharges that varied by product, with some products having higher and statistically significant overcharges, while others were lower and not statistically significant. Products with higher overcharges experienced greater overcharge volatility than those with lower overcharges, possibly because higher overcharges create scope for cheating. The size of overcharges on different products was affected by the assumptions made about the nature of the transition between collusion and competition.

A multi-product cartel that lives through multiple collusive phases will experience variations in overcharges over time. This confirms that differences in market conditions during different phases affect the ability of the cartel to raise prices to the same degree. Overcharges are likely to be lower during periods of economic slowdown. In the case of the wire cartel, overcharges

in the second phase were higher than those in the third phase. There were several potential reasons for this. A first reason was that the third phase emerged out of a price war, and the cartel was trying to re-establish itself. A second reason was that significant cost increases were experienced in the third phase, together with emerging from a price war in the face of an economic slowdown. It is possible that the cartel did not manage to fully exert itself in the third phase. A third, this phase coincided with greater cartel enforcement, which might have altered the cartel's price path.

The variability of overcharges on different products and over collusive phases has implications for cartel harm, penalties and damages proceedings, which are still rare in South Africa. Harm and damages must not be generalised through averages over different products and over time, without considering product-specific and phase-specific overcharges, especially where there is evidence that the cartel experienced instability and changing market conditions.

The dissertation also shows that the way a cartel transitions between collusion and competition affects overcharge estimates, meaning that estimation methods need to be adapted or equipped to deal with different forms of transition.

5.1.3 Chapter 4: A critical examination of South Africa's 20-year cartel enforcement record

The findings in Chapters 2 and 3 provided a case for a critical examination of the efficacy of the cartel enforcement record over the 20 years since the inception of the modern competition law regime. This was done in a comprehensive way to include an assessment of how effectively cartel enforcement tools have been used to detect and deter collusion. As is widely accepted, the author found that penalties and leniency together with settlements have driven cartel enforcement in terms of detection and prosecution. They were augmented by increased levels of funding over the years, which have increased the Commission's capacity.

Leniency and settlements have, however, not resulted in resource and time savings, which the Commission could have diverted to detecting other cartels. The duration of investigating and prosecuting firms has progressively increased over time. This is contrary to the predictions that underlie leniency and settlements, especially considering that over 90 per cent of prosecutions are resolved through settlements, and leniency. Econometric analysis of the relationship between the duration and key drivers of enforcement activity do not show a statistically significant negative relationship as expected, and where the relationship was as expected, the

results were not statistically significant. The dissertation categorises the sources of the prolonged investigation and prosecution of firms into four groups: first, an increased case load and complexity of cases; second, inefficiencies on the part of the Commission in case handling; third, strategic delays by firms in order to pay lower real penalties; and fourth, firms waiting to assess the strength of cases before committing to settlement decisions. Consequently, many firms are penalised only after a significant period has passed.

The dissertation posits that significant delays in finalising investigations and prosecuting firms has resulted in defendants paying heavily discounted penalties. This supports the view that firms have the incentive to delay payments as long as possible, since this reduces the size of the real penalties they eventually pay. The benefit of delays in paying penalties seems to outweigh the benefit of lower penalties due to the early settlement of cases. These delays are likely to undermine the deterrence effects of penalties.

The next section summarises the contributions of this dissertation.

5.2   Summary of contributions

The contributions made by this dissertation can be split into empirical and policy contributions, with the former discussed in section 5.2.1 and the latter in section 5.2.2.

5.2.1 Empirical contributions

This dissertation contributes to the body of empirical literature on collusion and cartel enforcement by presenting a systematic study of cartels and cartel enforcement (as presented in Chapter 2 and 4) and presenting a new method for the estimation of overcharges in a multi-product setting (in Chapter 3). The systematic analysis of cartels and cartel enforcement in Chapters 2 and 4, follows the methodology of the relevant international literature, including the well-known work by Porter (2005) and Levenstein and Suslow (2006) as discussed. Specifically, these chapters entailed comprehensively discussing the current industrial organization and competition policy literature on aspects of cartels and cartel enforcement. Therefore, the selection of characteristics, as well as the case study method employed, has a sound basis in economic theory and the related policy literature with several policy implications stemming from the study.

155

Chapter 2 shows that collusion in South Africa was extensive and durable, lasting several years on average. Most of the institutional features of the key South African cartels studied were consistent with those identified in the economic literature. Cartels use various mechanisms to manage internal and external threats. The important salient features observed in this dissertation relate to communication, mechanisms for rent sharing, the vertical features of collusion and the strategic use of trade tariffs.

The key South African cartels used a combination of different forms of communication and monitoring. The sales team infrastructure and relationships with customers can provide information on cheating by cartel members. This infrastructure does not require cartel-specific investments and, as such, is costless. Further, a cartel and its individual members do not need to centralise every aspect of their market intelligence to monitor cheating. These aspects are complementary and reduce the uncertainty created by imperfect centralised information. They are augmented by the learning effects derived from information exchange and repeated interaction.

Economic theory should incorporate the impact of complementary forms of communication if it is to advance to explaining collusion in instances where cartels face imperfect information. The uncertainties faced by cartels may be lower than the theory predicts. It is important to incorporate into the theory, the complementarities among the different forms of communication and monitoring that cartel members can use, and how these reduce uncertainty to sustain collusion. This is likely to advance the theory beyond focusing on the imperfections of cartel-level information. Admittedly, this may vary from cartel to cartel depending on each cartel's circumstances. Uncertainty is a permanent feature of most businesses, and firms learn to operate under such conditions. Collusion is one area where cartel members must adapt and manage the uncertainty to tolerable levels. Communication and monitoring are areas where the competition authorities could help further develop knowledge of collusion by fully investigating and reporting, in their public decisions, the various components of communication and monitoring that happen in cartels. This should be the case even if a case is settled.

The dissertation highlights the vertical nature of some of the major cartels in South Africa and points to the contribution that this feature makes to cartel stability and harm. The dissertation shows that the vertical structure of cartels stabilises collusion and enhances the ability of the cartel to exert pricing power along the value chain. In the cited examples, the vertical structure

appears to have helped align the interests of cartel members and to have addressed external threats from downstream and upstream firms that may have destabilised the collusive structure as they grew. It is important to account for the vertical effects when examining cartel harm.

The contribution of Chapter 2 to understanding especially the institutional features of cartels fits well into the international literature that recognises variation in cartel behaviour. It identifies factors that may well manifest in other developing-country and indeed developed country contexts. Further, the nature of the study is similar to that of leading published and widely cited studies on collusion such as Porter (2005) who studied organisational features of cartels and Levenstein and Suslow (2006) who examined the determinants of cartel success. The study in Chapter 2 is rooted in economics literature and reveals some factors which are often not directly/fully incorporated in economics theory and in studies of collusion. For example, the use of complementary communication methods, which might explain how cartels sustain collusion when certain aspects of monitoring are not centralised by the cartel. The study calls for further theoretical work to incorporate some of the findings. It is my view that this will contribute to the advancement of economics theory on collusion and help better understand collusion.

The empirical methods in Chapter 3 are new to the literature on cartel overcharges and damages. As the author explained in that Chapter, the panel model used is capable of estimating multi-product cartel overcharges. This is highly relevant in circumstances where a significant number of cartels involve multi-product collusion. The theoretical basis for the overcharges model used in Chapter 3 is consistent with the nature of cartel overcharges models used in the international literature, which mostly take a reduced-form specification. The dissertation shows that a multi-product cartel will exert itself in a variable way, by product and over different collusive phases. Variation of overcharges on different products and over time in a multi-product setting might reflect a cartel optimising cartel pricing across a portfolio of products or markets and adapting to temporal market conditions. This is not something that is often dealt with in most empirical studies of overcharges.

As far as Chapter 4 is concerned, a South African focus is a particular benefit in terms of contribution. South Africa is viewed favourably internationally among its developing-country peers in terms of competition law enforcement, including active cartel enforcement. As a result, a review of the South African experience potentially contributes to improvements in the efficacy of cartel enforcement, here in South Africa and in other jurisdictions. It is important

to also note that most jurisdictions follow the same cartel enforcement paths in terms of markets investigated with some multi-national companies participating in cartels in multiple jurisdictions and enforcement tools used (leniency, settlement procedures and penalties). The challenges identified and examined in this dissertation relating to the South African cartel enforcement record are therefore not unique to South Africa. For example, heavy-reliance on leniency, the inability of leniency to generate savings in time and resources for the authorities and the impact of delays in imposing penalties on deterrence also apply to jurisdictions such as the EU and the US. As such, the contribution to the efficacy of cartel enforcement has potential to apply in other jurisdictions.

The theoretical and empirical literature on the use of penalties for cartel deterrence could be developed further to explicitly account for the impact that delays have on deterrence. This dissertation shows that delays in the imposition and payment of penalties has a significant downward impact on the present value of the penalties paid. Firms appear to have made a trade-off between paying a lower negotiated penalty with early settlement (close to the beginning of the investigation) versus paying a lower discounted penalty resulting from delays in the finalisation of cases. South African firms routinely pay penalties after a prolonged lapse in time since the beginning of the investigation, even where there is leniency, and most cases are resolved through settlements.

More generally, theoretical advances are typically based on new insights emerging in the descriptive literature. For example, Chapter 2 which deals with the characteristics and institutional features of collusion in South Africa points to areas for future theoretical and empirical research. One such area is collusion under conditions of imperfect communication where cartels use complementary methods of communication and where some methods are less formalised. Such findings can contribute to the development of theory by providing theoretical IO scholars with insights that they can incorporate in existing or new game-theoretic models of cartel behaviour. This in turn then contributes to advancements in economics theory and empirical research.

It is worth mentioning that most studies on collusion that have been published in international peer-reviewed economic journals are based on cases studies of cartels that occurred in particular jurisdictions. Examples of this include Morton (1997) who studied entry deterrence by British cartels; Levenstein and Suslow (2011); and Hüschelrath, Müller and Veith (2013) who studied overcharges from the German cement cartel. As far as Chapter 3 of this dissertation

158

is concerned, published contributions related to cartel overcharges and damages estimation typically relate to specific case studies (see example of the German cement cartel). In addition, applied or empirical research on competition policy, based on South African case studies, has been published in several international journals over the past 15 years. For example, Khumalo, Mashiane and Roberts (2014) who study overcharges from the South African pre-cast concrete pipes cartel; Mncube (2013) who studied strategic entry deterrence in the bread cartel; Mncube (2014) who studied overcharges from the South African wheat flour cartel; Boshoff and van Jaarsveld (2019) who study recurrent collusion in the South African cement cartel. This dissertation adds to this body of empirical research, contributing with the use of techniques and producing findings that are not in these previous studies.

5.2.2 Policy contributions

There are several policy contributions made by the dissertation. The first policy contribution is that the future of effective cartel deterrence may lie in the ability of competition authorities to understand the characteristics and institutional features of cartels, and using these, to screen for, or gather forensic evidence of collusion.

The second policy contribution made by the dissertation is that informal communication may well be supported by the nature of the relationships between employees of different firms. Investigating communication in cartels may require the use of more forensic tools.

The third policy contribution in terms of policy is that competition authorities should examine the interplay between cartels at different levels, especially where firms take part in cartels at different levels, rather than treating cartels at each level of the value chain as separate infringements. Treatment of the cartels as being separate is likely to yield only partial appreciation of the full scope of collusion and its impact, leading to under-deterrence.

Fourth in terms of policy contributions made by the dissertation is that cartel investigations, or screening for collusion, should also entail searching for the existence of joint ventures, the incidence of sub-contracting, cross-payments, cross-supply, or toll-manufacturing arrangements between firms suspected of collusion. This is especially the case in those markets where contracts with customers are tender-based. Forensic investigation of tendering behaviour across space and time thus becomes important in cartel detection.

The fifth policy contribution is that multi-product firms suspected of collusion in one market should be screened for collusion in other markets, and the links between collusion in various markets should be examined because the management of collusion likely takes a quid pro quo structure that spans more than one market and product.

Sixth in terms of policy contributions is that in markets where imports present the next best alternative source of supply of key inputs – other than the firms suspected of cartel conduct – investigations of collusion should also incorporate an examination of the activities of cartel members and their trade bodies in advocating for or supporting the imposition of trade tariffs on competing imported products. This matters for understanding how a cartel deals with external threats, and it is also important with respect to consideration and design of remedies.

The seventh policy contribution is that given the variability of overcharges by product and over time, the right approach when determining penalties and damages should account for these two aspects, rather than assuming overcharge constancy along the two dimensions.

The eighth policy contribution – given delays and the prolonged nature of the investigations and prosecutions of firms – is that an optimal cartel enforcement policy would be one that considers the discount rates on penalties resulting from delays, and that uses sanctions in a manner that dynamically incentivises the early settlement, or finalisation, of cases by progressively and more aggressively penalising delays in settling cases. Further, competition authorities should consider adding interest to the penalty amount for the period between the year the penalty is calculated and when it is paid. This will enhance and preserve the deterrence effects of penalties.

The next section highlights suggested areas for future research.

5.3 Future research

The dissertation identifies several areas of future research. The first area of potential future research is the need to further examine, at both theoretical and empirical levels, the drivers and determinants of firm incentives to encourage or engage in collusion at more than one level of the value chain, given the impact of exercising market power at one level on demand and costs.

The second area of future research, related to the first area, regards the need to develop a framework for examining cartel overcharges where cartels have vertical features to them. Typically, cartel overcharges are studied at the level of the value chain where each cartel is

160

said to exist, with little or no attention paid to the vertical aspects of collusion. This requires both theoretical and empirical consideration, since it will involve examining the interplay between upstream and downstream markets and considering the trade-offs involved, which are likely to affect overcharges. Related to harm from collusion is the potential need to examine whether reliance on indirect communication affects the size of overcharges achieved. Related to harm from collusion is the potential need to examine whether reliance on indirect communication affects the size of overcharges achieved such cartels and whether these differ significantly from those imposed by cartels characterised by more direct communication.

In the third area of potential future research, any prospective future examination of cartel dynamics and estimation of cartel overcharges in markets where imports are an important source of potential constraints on various products should more explicitly consider the strategic use of trade policy tools by cartels to restrict the threat of external destabilising factors and to maintain supra-competitive prices. This will mean that prospective future studies should examine how the cartel price path is affected by trade policy tools. A further related area for future study relates to the willingness and role of government, whether inadvertently or as a result of trying to achieve other policy objectives such as supporting domestic industry. This should include how governments can balance legitimate need for policy intervention in certain markets against the risks to the competitive process and the impact on consumers and the economy.

A fourth area of potential future research is empirical and theoretical work on the optimum design and implementation of leniency, settlements, and penalty policies where firms strategically delay finalising cases and where the competition authority inefficiently handles cases. Both undermine deterrence. This will entail also examining the key drivers of the timing of firms' decisions to settle cases with the competition authorities, to account for the benefits that firms get from delaying cases. There is significant emphasis on firm incentives to settle early, appearing as though this is the dominant strategy. The cartel enforcement record shows that the dominant strategy of firms in South Africa may be to strategically delay settlement as long as possible, even where there is leniency. Empirical studies of firm sensitivity to penalties that get progressively higher with delays in settlement will contribute to better cartel enforcement policy.

This dissertation is a response to the need to examine collusion in South Africa in terms of its characteristics, and the institutional features of selected key cartels, with a view to

understanding the institutional factors contributing to its pervasiveness and durability. In addition, the dissertation was aimed at examining the harm resulting from a multi-product cartel that had lived through different collusive phases. Having identified the key institutional features of selected key South African cartels, including those that warranted emphasis, and having shown that multi-product cartels impose significant overcharges, the next area for examination was the cartel enforcement record. This resulted in a need to rethink the Commission's approach to penalties. This is important in terms of deterrence, given that leniency – which has hitherto been a key driver of cartel detection – has declined, and penalties remain the main tool for achieving deterrence. When the probability of detection declines, deterrence requires an increase in the level of penalties. At present, the deterring effects of penalties have been undermined by prolonged investigations and finalisation of cases, which have resulted in firms paying heavily discounted real penalties. Adding interest to the penalty amount for the period between the year the penalty is calculated and when the penalty is paid could help mitigate the negative impact of delays on penalties as a deterrence tool. This dissertation proposes an optimal cartel enforcement framework that incentivises the early conclusion of cases by progressively penalising firms for delays.

REFERENCES

Page, William H. 2010. Proving antitrust damages: Legal and economic issues. Chicago, Ill: *Section of Antitrust Law, American Bar Association*. Second edition. Washington D.C.

Abreu, D., Pearce, D., and Stacchetti, E. 1986. Optimal cartel equilibria with imperfect monitoring. *Journal of Economic Theory*, 39(1):251-269.

Albæk, S., Møllgaard, P., and Overgaard, P.B. 1997. Government-assisted oligopoly coordination? A concrete case. *The Journal of Industrial Economics*, 45(4):429-43.

Ascione, A., and Motta, M. 2010. Settlements in cartel cases, in Ehlermann, C.D., and Marquis, M. European Competition Law Annual 2008: Antitrust Settlements under EC Competition Law. *European Competition Law Annual*. Bloomsbury Publishing.

Athey, S., and Bagwell, K. 2001. Optimal collusion with private information. *The RAND Journal of Economics*, 32(3):428-65.

Athey, S., Bagwell, K., and Sanchirico, C. 2004. Collusion and price rigidity. *The Review of Economic Studies*, 71(2):317-49.

Bai, J., and Perron, P. 1998. Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47-78.

Bai, J., and Perron, P. 2003. Critical values for multiple structural change tests. *The Econometrics Journal*, 6(1):72-78.

Baker, J. 2002. Mavericks, mergers, and exclusion: Proving coordinated competitive effects under the antitrust laws. *New York University Law Review*, 77.

Bar-Gill, O., and Harel, A. 2001. Crime rates and expected sanctions: The economics of deterrence revisited. *The Journal of Legal Studies*, 30(2):485-501.

Bebchuk, L.A., and Kaplow, L. 1992. Optimal sanctions when individuals are imperfectly informed about the probability of apprehension. *The Journal of Legal Studies*, 21(2):365-370.

Bebchuk, L.A., and Kaplow, L. 1993. Optimal sanctions and differences in individuals' likelihood of avoiding detection. *International Review of Law and Economics*, 13:217-224.

Becker, G. 1968. Crime and punishment: An economic approach. *Journal of Political Economy*, 76.

Bernheim, B., and Whinston, M. 1990. Multimarket contact and collusive behavior. *The RAND Journal of Economics*, 21(1):1-26.

Benoit, J-P., and Krishna, V. 1985. Finitely repeated games. *Econometrica*, 53(4):905-22.

Binder, M., Hsiao, C., and Pesaran, M. H. 2005. Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory*, 21:795-837.

Biørn, E. 1992. Econometrics of panel data with measurement errors, in L. M´aty´as and P. Sevestre. eds. The econometrics of panel data: Theory and applications. Kluwer Academic Publishers, Dordrecht, 1st ed. 152-195.

Bishop, S., and Walker, M. 2010. The economics of EC competition law: Concepts, application and measurement. Sweet & Maxwell.

Block, M.K., Nold, F.C., and Sidak, J.G. 1981. The deterrent effect of antitrust enforcement. *Journal of Political Economy*, 89(3):429-45.

Blundell, R., and Bond, S. 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*. **87** (1): 115–143.

Bork, R.H. 1978. *The antitrust paradox: a policy at war with itself*. New York: Free Press, 50. ISBN 0-465-00369-9.

Boshoff, W.H. 2015. Illegal cartel overcharges in markets with a legal cartel history: Bitumen prices in South Africa. *South African Journal of Economics*, 83(2):220-239.

Boshoff, W.H., and Van Jaarsveld, R. 2019. Recurrent collusion: Cartel episodes and overcharges in the South African cement market. *Rev Ind Organ*, 54:353-380.

Boswijk H.P., Bun, M.J.G., and Schinkel, M.P. 2019. Cartel dating. *Journal of Applied Economics*, 34:26-42.

Breitung, J. 2000. The local power of some unit root tests for panel data, in Advances in Econometrics, Vol. 15: Nonstationary panels, panel cointegration, and dynamic panels, ed. B. H. Baltagi, 161-178. Amsterdam: JAI Press.

Burns, M.R. 1986. Predatory pricing and the acquisition cost of competitors. *Journal of Political Economy*, 94(2):266-96.

Cave, J., and Salant, S. 1987. Cartels that vote: Agricultural marketing boards and induced voting behavior. Centre for Research on Economic and Social Theory, Crest Working Paper 87(30).

Choi, I. 2001. Unit root tests for panel data. *Journal of International Money and Finance*, 20:249-272.

Ciliberto, F., and Williams, J.W. 2014. Does multimarket contact facilitate tacit collusion? Inference on conduct parameters in the airline industry. *The RAND Journal of Economics*, 45:764-791.

Clay, K., and Werner T. 2002. Strategic behavior in whiskey distilling, 1887-1895. *The Journal of Economic History*, 62(4):999-1023.

Connor, J.M., and Lande, R. 2006. The size of cartel overcharges. *Antitrust Bulletin*, 51:983-1022.

Connor, J.M. 2008. Forensic economics: An introduction with special emphasis on price fixing. *Journal of Competition Law and Economics* 4:31-59.

Connor, J.M. 2014. Cartel overcharges. *Research in Law and Economics*, 26:249-387.

Connor, J.M. 2017. Cartels costly for customers. *SSRN Scholarly Paper*. Rochester, NY: Social Science Research Network, 18 June. https://papers.ssrn.com/abstract=2988489.

Crede, C.J. 2019. A structural break cartel screen for dating and detecting collusion. *Rev Ind Organ*, 54:543-574.

De Coninck, R. 2010. Application of the nonhorizontal merger guidelines. *Antitrust Bulletin*, 55:929-952.

Dick, A.R. 1996. Identifying contracts, combinations and conspiracies in restraint of trade. *Managerial and Decision Economics*, 17(2):203-16.

Dixit, A. 1979. Model of duopoly suggesting a theory of entry barriers. *A.J. Reprints Antitrust L. & Econ*, 10:399.

Eckbo, P.L. 1976. The future of world oil. Cambridge: Ballinger Publishing. U.S.

Edwards, K., and Padilla, J.A. 2010. Antitrust settlements in the EU: Private incentives and enforcement policy, in Ehlermann, C.D., and Marquis, M. European competition law annual 2008: Antitrust settlements under EC competition law. *European Competition Law Annual*. Bloomsbury Publishing.

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

Frank, N., and Schliffke, P. 2013. The post-cartel equilibrium puzzle in the German cement market: A reply to Hüschelrath, Müller, and Veith. *Journal of Competition Law and Economics*, 9:495-509.

Franzoni, L. 1999. Negotiated enforcement and credible deterrence. *Economic Journal*, 109(458):509-535.

Friedman, J.W. 1971. A non-cooperative equilibrium for supergames. *The Review of Economic Studies*, 38(1):1-12.

Genesove, D., and Mullin, W.P. 2001. Rules, communication, and collusion: Narrative evidence from the Sugar Institute case. *The American Economic Review*, 91(3):379-98.

Ghosal, V. 2011. Regime shift in antitrust laws, economics and enforcement. *SSRN Scholarly Paper*. Rochester New York: Social Science Research Network. https://papers.ssrn.com/abstract=1020448.

Ghosal, V., and Sokol, D.D. 2016. Policy innovations, political preferences, and cartel prosecutions, *Rev. Indust. ORG.*, 48:405.

———. 2018. The rise and (potential) fall of U.S. cartel enforcement, *SSRN Scholarly Paper*. Rochester New York: Social Science Research Network. https://papers.ssrn.com/abstract=3162867.

Granger, C.W. J. 1990. Aggregation of time-series variables: A survey, in T. Barker and M. H. Pesaran. eds. Disaggregation in econometric modeling. Routledge London.

Green, E.J., and Porter, R.H. 1984. Noncooperative collusion under imperfect price information. *Econometrica*, 52(1):87-100.

Grether, D.M., and Plot, C.R. 1984. The effects of market practices in oligopolistic markets: An experimental examination of the Ethyl case. *Economic Inquiry*, 22:479-507.

Griliches, Z., and Hausman, J. A. 1986. Errors-in-variables in panel data. *Journal of Econometrics*, 31:93-118.

Hadri, K. 2000. Testing for stationarity in heterogenous panel data. *Econometrics Journal*, 3:148-161.

Harrington Jr., J.E. 1987. Collusion in multi-product oligopoly games under a finite horizon. *International Economic Review*, 28(1):1-14.

———. 2004. Cartel pricing dynamics in the presence of an antitrust authority. *The RAND Journal of Economics*, 35(4):651-673.

———. 2004. Post-cartel pricing during litigation. *Journal of Industrial Economics*, 52:517-533.

———. 2015. Some thoughts on why certain markets are more susceptible to collusion. *OECD Global Forum on Competition – "Serial Offenders"*. available at: https://www.oecd.org/competition/globalforum/Harrington_OECD_10%2015.pdf.

———. 2015. The current state of the theory of collusion: unexplained phenomena and unexplored directions. Bergen Competition Policy Conference, Bergen Centre for Competition Law and Economics. 23-24 April.

———. 2016. Successes and challenges in the fight against cartels. Pre-BRICS International Competition Conference (November 2015) available at. https://joeharrington5201922.github.io/pdf/Pre-BRICS_Harrington.pdf.

Harrington, J.E., and Chang, M. 2009. Modeling the birth and death of cartels with an application to evaluating competition policy. *The Journal of the European Economic Association*, 7:1400-1435.

———. 2015. When Can We Expect a Corporate Leniency Program to Result in Fewer Cartels? *The Journal of Law & Economics*, 58(2):417-449.

Harris, R. D. F., and Tzavalis, E. 1999. Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of Econometrics*, 91:201-226.

Hellwig, M., and Hüschelrath, K. 2016. Cartel cases and the cartel enforcement process in the European Union from 2001 to 2015: A quantitative assessment, *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2848862.

Hendricks, K., Porter, R., and Tan, G. 2008. Bidding rings and the winner's curse. *The RAND Journal of Economics*, 39(4):1018-1041.

Higgins, R.S., and Johnson, P.A. 2003. The mean effect of structural change on the dependent variable is accurately measured by the intercept change alone. *Economics Letters*, 80:255-259.

Holt, C.A., and Scheffman, D.T. 1987. Facilitating practices: The effects of advance notice and best-price policies. *The RAND Journal of Economics*, 18(2):187-97.

Hsiao, C. 2005. Why panel data? *Singapore Economic Review*, 50(2):1-12.

Hsiao, C. 2006. Longitudinal data analysis, in the New Palgrave Dictionary of Economics. Palgrave MacMillan Press.

Hsiao, C., Appelbe, T. W., and Dineen, C. R. 1993. A general framework for panel data analysis - with an application to Canadian customer dialed long distance service. *Journal of Econometrics*, 59:63-86.

Hsiao, C., Chan, M. W., Mountain, D. C., and Tsui, K. Y. 1989. Modeling Ontario regional electricity system demand using a mixed fixed and random coefficients approach. *Regional Science and Urban Economics*, 19:567-587.

Hsiao, C., Mountain, D. C., and Ho-Illman, K. 1995. Bayesian integration of end-use metering and conditional demand analysis. *Journal of Business and Economic Statistics*, 13:315-326.

Hsiao, C., Shen, Y., and Fujiki, H. 2005. Aggregate vs disaggregate data analysis - a paradox in the estimation of money demand function of Japan under the low interest rate policy. *Journal of Applied Econometrics*, 20:579–601.

Hüschelrath, K., Müller, K., and Veith, T. 2013. Concrete shoes for competition: The effect of the German cement cartel on market price. *Journal of Competition Law and Economics*, 9:97-123.

Hüschelrath, K., & Veith, T. 2013. Cartel detection in procurement markets. *Managerial and Decision Economics*, 35(6), 404-422.

Im, K. S., Pesaran, M. H., and Shin, Y. 2003. Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, 115:53-74.

Jacquemin, A., Nambu, T., and Dewez, I. 1981. A dynamic analysis of export cartels: The Japanese case. *Economic Journal*, 91(363):685-96.

Kaplow, L. 1989. The optimal probability and magnitude of fines for acts that are definitely undesirable. *National Bureau of Economic Research*, Working Paper No. 3008. http://www.nber.org/papers/w3008.

Katsoulacos Y., Ulph D. 2013. Antitrust penalties and the implications of empirical evidence on cartel overcharges. *The Economic Journal*, 123:558-581.

Katsoulacos, Y., Motchenkova, E., and Ulph, D. 2015. Penalizing cartels: The case for basing penalties on price overcharge. *International Journal of Industrial Organization*, 42:70-80.

Katsoulacos, Y., Motchenkova, E., and Ulph, D. 2017. Penalizing on the basis of the severity of the offence: A sophisticated revenue-based cartel penalty. *TILEC Discussion Paper* No. 2017-014, Available at:
SSRN: https://ssrn.com/abstract=2970395 or http://dx.doi.org/10.2139/ssrn.2970395.

Khumalo, J., Mashiane, J., and Roberts, S. 2014. Harm and overcharge in the South African precast concrete products cartel. *Journal of Competition Law and Economics*, Oxford University Press, 10(3):621-646.

Kovacic, W.E., Marshall, R.C., and Meurer, M.J. 2018. Serial collusion by multi-product firms, *SSRN Scholarly Paper*. Rochester, NY: Social Science Research Network. 13 August. https://papers.ssrn.com/abstract=3235398.

Landes, W.M. 1983. Optimal sanctions for antitrust violations. *The University of Chicago Law Review*, 50(2):652.

Levenstein, M. 1995. Mass production conquers the pool: Firm organization and the nature of competition in the nineteenth century. *The Journal of Economic History*, 55(3):575-611.

Levenstein, M., and Suslow, V.Y., 2004. Contemporary international cartels and developing countries: Economic effects and implications for competition policy. *Antitrust Law Journal*, 71(3):801-52.

———. 2006. What determines cartel success? *Journal of Economic Literature*, 44(1):43-95.

———. 2011. Breaking up is hard to do: Determinants of cartel duration. *The Journal of Law and Economics*, 54(2):455-92.

Levin, A., Lin, C., and Chu, J. 2002. Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics*, 108:1-24.

Matsushima, H. 2001. Multimarket contact, imperfect monitoring, and implicit collusion. *Journal of Economic Theory*, 98(1):158-178.

Mncube, L. 2013. Strategic entry deterrence: Pioneer Foods and the bread cartel. *Journal of Competition Law and Economics*, 9:637-654.

Mncube, L. 2014. The South African wheat flour cartel: overcharges at the mill. *Journal of Industry, Competition and Trade*, 14:487-509.

Mondliwa, P., and Das Nair, R. 2017. Overcharge Estimates in the South African Reinforcing Bar Cartel. *CCRED Working Pape*r No. 2017/6 available at:
SSRN: https://ssrn.com/abstract=2947581 or http://dx.doi.org/10.2139/ssrn.2947581.

Morton, F.S. 1997. Entry and Predation: British Shipping Cartels 1879-1929. *Journal of Economics & Management Strategy*, 6:679-724.

Motta, M. 2004. Competition Policy: Theory and Practice. New York: Cambridge University Press.

Motta, M. 2008. On cartel deterrence and fines in the European Union. *E.C.L.R.*, 209.

Nerlove, M. 2002. Essays in Panel Data Econometrics. Cambridge University Press. Cambridge.

Nieberding, J.F. 2006. Estimating overcharges in antitrust cases using a reduced-form approach: methods and issues. *Journal of Applied Economics*, 9(2):361-380.

Niels, G., Jenkins, H., and Kavanagh, J. 2011. Economics for competition lawyers. Oxford University Press. New York.

O'Donoghue R., and Padilla A.J. 2013. The law and economics of Article 102 TFEU. 2nd Edition. Hart Publishing. Oxford & Portland Oregon.

OECD, 2002, Report on the nature and impact of hard core cartels and sanctions against cartels under national competition laws. 9 April.

Oxera et al. 2009. Quantifying antitrust damages: Towards non-binding guidance for courts. report prepared for DG Competition, December.

Paha, J. 2011. Empirical methods in the analysis of collusion. *Empirica*, 38:389-415.

Pakes, A., and Griliches, Z. 1984. Estimating distributed lags in short panels with an application to the specification of depreciation patterns and capital stock constructs. *Review of Economic Studies*, 51:243-262.

Persyn, D., and Westerlund, J. 2008. Error correction based cointegration tests for panel data. *Stata Journal*, 8(2):232-241.

Pesaran, M. H. 2003. On aggregation of linear dynamic models: An application to life-cycle consumption models under habit formation. *Economic Modeling*, 20:227-435.

Pesaran, M. H., and Smith, R. 1995. Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 68:79-113.

Phillips, P. C. B., and Moon, H. R. 1999. Linear regression limit theory for nonstationary panel data. *Econometrica*, 67:1057-1111.

Podolny, J.M., and Scott Morton, F. 1999. Social status, entry and predation: The case of British shipping cartels 1879-1929. *The Journal of Industrial Economics*, 47(1):41-67.

Porter, R.H. 2005. Detecting collusion. *Review of Industrial Organization*, 26(2):147-67.

Posner, R.A. 1970. A statistical study of antitrust enforcement. *The Journal of Law and Economics*, 13(2):365-419.

———. 1976. Antitrust law: An economic perspective. University of Chicago Press.

———. 2001. Antitrust law. 2nd edition. Chicago: University of Chicago Press.

Ratshisusu, H. 2014. Limiting collusion in the construction industry: A review of the bid-rigging settlement in South Africa. *Journal of Economic and Financial Sciences*, 7(1):587-606.

Rotemberg, J.J., and Saloner, G. 1986. A supergame-theoretic model of price wars during booms. *American Economic Review*, 6(3):390-407.

Rubinfeld, D. L. 1985. Econometrics in the courtroom. *Columbia Law Review*, 85:1048-1097.

Spagnolo, G. 1999. On interdependent supergames: multimarket contact, concavity, and collusion. *Journal of Economic Theory*, 89(1):127-139.

Spagnolo, G., and Marvão, C.M.P. 2016. Cartels and leniency: Taking stock of what we learnt. *SSRN Scholarly Paper*. Rochester New York: Social Science Research Network. 9 September. https://papers.ssrn.com/abstract=2850498.

Stigler, G.J. 1964. A theory of oligopoly. *The Journal of Political Economy*, 72(1):44-61.

Symeonidis, G. 2003. In which industries is collusion more likely? Evidence from the UK. *The Journal of Industrial Economics*, 51(1):45-74.

Van Dijk, T., and Verboven, F. 2008. Quantification of damages, in Collins, W.D. (Hrsg): Issues in Competition Law and Policy. *ABA Publications in Antitrust*.

Wansbeek, T. J., and Koning, R. H. 1989. Measurement error and panel data. *Statistica Neerlandica*, 45:85-92.

Warden, G. 2003. The effect of antitrust policy on consumer welfare: What Crandall and Winston overlook. *Economic Analysis Group discussion paper*, January.

Weiman, D.F., and Levin, R.C. 1994. Preying for monopoly? The case of Southern Bell Telephone Company, 1894-1912. *Journal of Political Economy*, 102(1):103-26.

White, H., Marshall, R., and Kennedy, P. 2006. The measurement of economic damages in antitrust civil litigation. *ABA Antitrust Section Economic Committee Newsletter*, 6:17-22.

Wils, W.P.J. 2010. The use of settlements in public antitrust enforcement: Objectives and principles, in Ehlermann, C.D., and Marquis, M., European Competition Law Annual 2008: Antitrust Settlements under EC Competition Law. *European Competition Law Annual*. Bloomsbury Publishing.

World Bank. 2016. South Africa Economic Update. Promoting Faster Growth and Poverty Alleviation Through Competition. Edition 8, New York.