e-book Overcharge

The Effect of U.S. v. Apple Inc. et. al. (2013) on e-book Prices

Isaac Anthony
Haverford College 2014

Abstract
This thesis sets out to estimate the effect of U.S. v. Apple Inc. et. al. (2013) on e-book prices by implementing a reduced form model, dummy variable approach (Nieberding, 2006), using a difference-in-differences analysis (Ashenfelter, 2013). This study follows from a deep history of relevant literature on constructing estimation of cartel overcharge models and using them to estimate but-for prices on real-world cartels (Connor, 2001; Clarke and Evenett, 2003; Frank and Schliffke, 2013). Here, I estimate the change in the accused publishers’ e-book retail prices relative to Random House (the one publisher among the Big Six never accused of collusion) due to each of the accused publisher's respective settlement with the U.S. Department of Justice and the disbanding of their agency contracts with their retailers. I find a 16.7 percent average decrease in e-book retail prices from the time-period immediately before to immediately after each publisher’s settlement date with the D.O.J. These findings suggest that e-book retail prices have been restored to their pre-collusion, competitive levels.
# Table of Contents

<table>
<thead>
<tr>
<th>Section (#)</th>
<th>Title</th>
<th>Page (#)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Background</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>Literature Review</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>Data</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>a. Primary Data Source</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>b. Secondary Data Source</td>
<td>22</td>
</tr>
<tr>
<td>5</td>
<td>Theory</td>
<td>23</td>
</tr>
<tr>
<td>6</td>
<td>Methods</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>a. Model I</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>b. Model II</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>c. Model III</td>
<td>35</td>
</tr>
<tr>
<td>7</td>
<td>Hypotheses</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>a. Hypothesis I</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>b. Hypothesis II</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>c. Hypothesis III</td>
<td>39</td>
</tr>
<tr>
<td>8</td>
<td>Results</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>a. Results for Hypothesis I</td>
<td>41</td>
</tr>
</tbody>
</table>
Special thanks to

Dan Lubart from iobyte Solutions, Inc. and

Professor Richard Ball.
Section 1: Introduction

This paper sets out to answer the research question: what was the effect of *U.S. v. Apple Inc. et. al. (2013)* on e-book prices? In a landmark case, the Department of Justice’s Antitrust Division ("D.O.J.") accused Apple Inc., AAPL, ("Apple") and five of the six largest e-book publishers, collectively referred to as the Big Six, of colluding to set the retail price of e-books in violation of Section 1 of the Sherman Antitrust Act (1890). The Big Six Publishers include Hachette Book Group Inc. ("Hachette"), HarperCollins Publishers L.L.C. ("HarperCollins"), Holtzbrinck Publishers L.L.C. ("Macmillan"), Penguin Group Inc. ("Penguin"), Random House L.L.C. ("Random House"), and Simon & Schuster Inc. ("Simon & Schuster"). The Big Six account for over 90 percent of e-books sold in the United States (Cote, 2013). Random House was the only publisher among them never accused of collusion with Apple.

In statements provided by the D.O.J. at trial, Apple allegedly helped to organize a massive price-fixing operation with the five accused publishers, allowing for these publishers to simultaneously shift their pricing from a wholesale model, where the publishers set the wholesale price of their e-books and the retailers set the corresponding retail price faced by customers, to an agency model of e-book distribution, where the publishers set the retail price and the retailers serve as their agents in selling e-books to consumers, collecting a fixed percentage of all sales. This new agency model effectively eliminated competition on price at the e-book retail level. Additionally, Apple’s contracts with the accused publishers included a most-favored-nation clause demanding that all five publishers enforce this agency model on all e-book retailers. As such, these agreements ensured that all other e-book retailers, notably Amazon.com, Inc., AMZN, ("Amazon") and Barnes & Noble, Inc., BKS, ("Barnes & Noble"), would have to match Apple’s prices by selling their e-books for the publisher-set prices.

By the date of the trial in June, 2013, five of the six accused publishers had settled with the D.O.J., so Apple was the only remaining defendant. Macmillan settled approximately one month after the trial. Section 1 of the Sherman Act bans “every contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or
commerce.” As can be seen in Figure 1, this law is rooted in industrial-organization economic theory because price fixing can allow producers to elevate prices above their marginal cost, leading to reduced consumer surplus and societal welfare (Pepall et. al., 2008). Section 1 of the Sherman Act makes price fixing per se illegal. This means that if the prosecution in an antitrust lawsuit can prove that price fixing has occurred, the defendants are guilty regardless of whether the fixed prices actually were higher than their competitive prices and no further inquiry is necessary. Per se price fixing is contrasted to a rule of reason approach when analyzing cases of price fixing, where the prosecution would need to provide credible estimates of the price increase due to the defendants’ price fixing for their actions to be considered illegal.

Evidence supplied at trial by the D.O.J. estimated a 16-19 percent across-the-board increase in e-book retail prices after the adoption of Apple’s agency contracts with the accused publishers (Ashenfelter, 2013). Representatives for Apple countered by arguing that the D.O.J. was targeting the wrong antitrust offender in the e-book publication and distribution industries as Amazon had an estimated 80 percent market-share of the e-book retail market as of Q1, 2010. Apple contended that their agency contracts with the accused publishers were necessary for their entrance into the e-book retail market, where they would compete with Amazon and help consumers face reduced e-book prices.

In her verdict, New York District Judge Denise Cote sided with the D.O.J. and on July 20, 2013 ruled Apple Inc. guilty of horizontal (within-industry) price fixing. The District Court determined that Apple's most-favored-nation clause defeated price competition (per se price fixing) and their agency contracts resulted in higher prices across the e-book retail industry (price fixing in violation of the rule of reason). After the trial, Apple was forced to abandon their agency contracts with all publishers.

My thesis investigates whether the prosecution of Apple by the D.O.J. and the verdict of U.S. v. Apple Inc. et. al. (2013) resulted in a subsequent decrease in the average e-book retail prices faced by consumers. A statistically-significant decrease in average e-book retail prices would suggest that the D.O.J.’s prosecution of the accused
publishers and Apple was an effective antitrust-policy intervention.

To calculate the change in e-book retail prices associated with the trial, I use an estimation of cartel overcharge methodology implementing a reduced form model, dummy variable approach (Nieberding, 2006), difference-in-differences analysis (Ashenfelter, 2013). A reduced form model, dummy variable approach compares what actually happened to prices in the affected industry with what would have happened “but for” the existence of a cartel. This model is one of the most popular for estimation of cartel overcharge analyses (Finkelstein and Levenbach, 1983), and has been used by much of the previous literature such as Connor's (2001) study of a global cartel of lysine manufacturers, Clarke and Evenett's (2003) study of an international cartel of vitamin manufacturers, and Frank and Schliffke's (2013) study of a German cartel of cement manufacturers. The variable of interest in the model is a dummy variable set to unity during the time of collusion and to zero otherwise. This variable’s coefficient measures the effect of the cartel on the cartelized good ceteris paribus.

To assist in their analysis on the anticompetitive effects of Apple’s agency contracts, members of the prosecution hired Orley Ashenfelter, the Joseph D. Green Professor of Economics at Princeton University, to serve as an economic consultant at trial. In his testimony, Ashenfelter (2013) provided an estimation of cartel overcharge using a reduced form method, dummy variable approach, difference-in-differences analysis. In this model, the change in the prices of goods sold by a cartel (or a group of cartels) as it transitions from a period of collusion to a period of competition (or vice versa) is compared to the changes in the prices of goods sold by a control group over the same time-period. Here, Random House L.L.C. is used as the control group, because Random House was the only publisher among the Big Six who never adopted agency contracts and thus was never accused of price collusion by the D.O.J. Ashenfelter (2013) estimated a 16.8 percent increase in the average e-book retail prices of the accused publishers due to their adoption of agency contracts relative to Random House.

My model almost directly mirrors Ashenfelter’s model except for the fact that he measures the change in the e-book retail prices for the conspiring publishers due to their
adoption of the agency pricing model, whereas my study measures the change in e-book retail prices due to *U.S. v. Apple Inc. et. al. (2013)* when these agency contracts were dissolved, e-book retailers regained some level of control over their prices, and price competition was reintroduced into the e-book retail industry. Not only do I look at a different time-period and dataset as Ashenfelter, but I also I add to the previous literature by extending his model to include multiple collusion dummy variables and time-periods.

For my data, I analyze weekly e-book prices extending 91 weeks (23 months) from April 8, 2012 to December 29, 2013 for a sample of 500 best-selling ISBNs for Amazon, Apple, and Barnes & Noble, the three largest firms in the e-book retail industry and for all six of the Big Six publishers. The data was received from iobyte Solutions, a consulting company for e-book publishers. Like Ashenfelter (2013), I develop a model including ISBN-level, retailer-level, and publisher-level fixed effects and conduct a difference-in-differences estimation by using ISBNs published by Random House as a control group. The difference-in-differences estimators for the five accused publishers estimate the change in the average e-book retail price due to the publishers settlements with the D.O.J. for ISBNs published by the accused publishers relative to the base-line ISBNs published by Random House. I also use a secondary dataset compiled using the Google Books A.P.I. This dataset allows me to construct an alternate model including ISBN-level controls such as e-book genre, page count, and age instead of ISBN-level fixed effects. This model serves as a test of robustness for my findings.

I find that the average e-book retail prices for the five accused publishers decreased from the time-period immediately before each publisher’s corresponding settlement date with the D.O.J. to the time-period immediately after relative to Random House over the same time-periods by 16.7 percent (significant at the one percent level). Additionally, I find that average e-book retail prices for all five of the accused publishers decreased by 7.8 percent (significant at the one percent level) relative to Random House over the entire time-period of my analysis.

These results can be compared to Ashenfelter’s (2013) findings of a 16.8 percent increase after the introduction of Apple’s agency contracts. My estimate of a 16.7 percent
average D.i.D. price-decrease from the time-period before each of the accused publisher’s settlement date to the time-period immediately after is insignificantly different from Ashenfelter’s estimate. This provides further evidence that post-agency e-book retail prices have returned to their competitive, pre-agency levels.

These estimates affirm my hypothesis: on average e-book retail prices were elevated during the collusion time-period, the period in which agency contracts were in effect and Apple was found guilty of price-fixing for the five accused publishers relative to Random House, the control group, and these prices fell after the accused publishers’ settlements with the D.O.J. These results add credence to Judge Cote’s decision by providing evidence of the anticompetitive effects of Apple's agency contracts with the five accused publishers. In addition, these results can be viewed as evidence refuting Apple’s continued plea of innocence, who on February 15, 2014 appealed Cote’s verdict.

The rest of this paper is structured as follows. Section 2 provides background information on U.S. v. Apple Inc. et. al. (2013). Section 3 presents a survey of related literature. Section 4 gives an overview of the data included in this analysis. Section 5 explains the theoretical background of my model. Section 6 lays out my methodology. Section 7 outlines my hypotheses. Section 8 summarizes all results, and Section 9 concludes.

Section 2: Background

An electronic book, or e-book, is a publication in digital format that is often the electronic version of a print book. Today, e-books may be read on a variety of devices including dedicated e-readers (such as Amazon’s Kindle or Barnes & Noble’s Nook), multipurpose tablets (such as Apple’s iPad), smartphones, and personal computers. According to the Association of American Publishers, as of Q1, 2013, the e-book market was worth almost $400 million, approximately 20 percent of the more general U.S. book market. The industry was characterized by a three-firm oligopoly with dominant market shares calculated as a percentage of e-book sales possessed by Amazon.com, Inc., Barnes & Noble, Inc., and Apple, Inc. Much of the background information surrounding the
case can be found in the *Opinion of the Court* delivered by New York District Judge Denise Cote (2013). Unless otherwise noted, the facts presented in this section can be attributed to this source (Cote, 2013). Additionally, a timeline of all of the events related to *U.S. v. Apple Inc. et. al. (2013)* as detailed in this section is presented in Figure 2.

On April 11, 2012, the Antitrust Division of the U.S. Department of Justice filed an antitrust lawsuit alleging that Apple and five of the six largest e-book publishing companies in the United States, known as the Big Six, had conspired to raise, fix, and stabilize the retail price for newly released and best-selling e-books in violation of Section 1 of the Sherman Act, which bans “every contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States” (Sherman Antitrust Act, 1890). Consequently, four of the five accused publishers settled with the D.O.J. prior to trial. Macmillan settled approximately one month after the date of the verdict. On June 3, 2013, Apple was brought to bench trial as the sole defendant in *U.S. v. Apple Inc. et. al. (2013)*, *U.S. District Court, Southern District of New York, No. 12-02826*. Much of the evidence for the trial included emails, taped conversations, and witness statements testifying to the facts presented below. Redacted versions of all of these documents can be found in the *Court Materials* provided online by the D.O.J.


Amazon's Kindle, launched in 2007, was the first e-reader to gain widespread commercial acceptance. Through 2009, Amazon dominated the e-book retail market, selling nearly 90 percent of all e-books. Barnes & Nobles' Nook was released two years later, in November of 2009, offering some competition to Amazon. The iPad was released in April, 2010. In 2009, the e-book market was valued at about $100 million, but was
observed to be growing exponentially. Estimates put the e-book industry’s growth at approximately 406 percent per year (Ashenfelter, 2013).

Prior to April, 2010, the Big Six publishers distributed e-books through a wholesale pricing model. The publisher for a particular ISBN set a list price (also referred to as a suggested retail price) and then sold the e-book to a retailer for a wholesale price, generally calculated at a 20 percent discount from the wholesale price of the corresponding print version of the e-book. The retailer then offered the e-book to consumers at whatever retail price it chose. Amazon’s business model implemented a discount-pricing strategy, generally guaranteeing a $9.99 price point for certain new release and best-selling e-books by discounting the publisher’s suggested retail price to determine their final retail price. According to representatives from the Big Six publishers, this $9.99 price point was estimated to be roughly the perfectly competitive price, matching the wholesale price for which Amazon purchased many of their e-books from them. Analysts speculated that Amazon's low e-book price guarantee was a method to drive consumers to purchase their Kindle e-reader with which they would then make many additional purchases in the Kindle App Store.

This low $9.99 price point for Amazon’s newly released e-books worried the Big Six publishers. They saw it as eating into the sales of their more profitable hardcover new releases, which were often priced at $30 or more. In this way, Amazon’s low e-book price point was part of their broader, multi-market profit maximization strategy, which included e-readers, apps, and other media distributed through their Kindle devices. For the publishers, on the other hand, Amazon’s low e-book retail prices were undermining the price of goods they offered in a different market: print books. Additionally, the publishers felt threatened by Amazon's growing power as the dominant e-book retailer. If Amazon solidified their hold on the e-book retail market, it was rumored that Amazon could potentially start using their monopsony bargaining power to negotiate reduced future e-book wholesale price points with the publishers. Or even worse, Amazon could start communicating directly with authors and literary agents for future contracts, cutting out the publishers entirely, such as has been done with their growing new service: Amazon Kindle.
Direct Publishing.

According to internal statements, in 2009 the Big Six tried to coordinate their efforts to put pressure on Amazon to abandon their $9.99 price point. In early 2009, the accused publishers jointly set the wholesale price for e-books at several dollars above $9.99, but Amazon continued selling new-release New York Times Bestsellers for $9.99, incurring losses on all sales. Later in 2009, the publishers tried adopting a strategy called “windowing” where they delayed release of the e-book version of new releases to try to mitigate the effect of lower e-book prices on hard-cover, new-release sales. Even after the company saw significantly reduced sales, Amazon’s $9.99 price point persisted. As of 2014, windowing is still in effect by e-book publishers.

Enter Apple. In late 2009, Apple was planning the date of release for their iPad as January 27, 2010. They hoped to coordinate their entry into the e-book market, symbolized by the launch of their new online e-book retail store, the iBookStore, with the launch of their new e-reader, the iPad. Apple representatives said they would do so only if they had contracts in hand with all six of the Big Six publishers guaranteeing their e-books would be supplied for the iPad. In mid-December, 2009, during the first meeting between Apple and representatives from the Big Six, Apple assured the publishers that they were willing to work with them to raise new-release e-book prices, suggesting $12.99 and $14.99 price points.

Over the course of their negotiations in December, 2009 and January, 2010, Apple and the publishers hoped to shift price competition from the retail level to the publisher level. Apple did not want to directly compete with Amazon (or any other e-book retailer) on price, and the accused publishers wanted to significantly increase the prevailing price point for new-release e-books. Apple offered the publishers the opportunity to move from a wholesale business model, where a publisher designates its wholesale price for each e-book and the retailer sets the retail price, to an agency model, where a publisher sets the final retail price and the retailer sells the e-book as its agent for a fixed percentage of the sale.

Under their newly proposed agency contracts with the publishers, Apple agreed to
receive a 30 percent agency fee for all sales through their iBookStore. The e-book retail prices, as set by the publishers, were floored into tiers as a percentage of their physical book's price point. To prevent publishers from deviating from the agreement, Apple's new contracts detailed a most-favored-nation clause, which imposed severe financial penalties upon any of the agreeing publishers for failure to enforce their agency pricing model on all other e-book retailers. This guaranteed that all e-book prices in Apple's iBookStore, as set by the publishers, matched all of Amazon's e-book prices in their Kindle App Store, which matched all other e-book retail prices across the industry.

By January 26, 2010, the week of the launch of Apple's iPad (January 21, 2010), five of the Big Six publishers had agreed to sign agency contracts with Apple. Random House refused. Apple decided to go ahead with the launch of their iPad anyway, and just as they expected, after the iBookstore opened in April, 2010, the price floors in their contracts became the new retail prices for the accused publishers' e-books. In the five months that followed, the accused publishers collectively priced 86 percent of their new-release titles sold through Amazon and 92 percent of their new-release titles sold through Apple within 1 percent of their designated price floors. This was also true for 99 percent of the New York Times Bestseller titles on Apple’s iBookstore, and 97 percent of New York Times Bestsellers sold through Amazon. The price increases at Amazon within roughly two weeks of moving to the agency model amounted to an average per unit e-book retail price increase of 14 percent for their new releases, 43 percent for their New York Times Bestsellers, and 19 percent across all of the accused publishers’ e-books (Cote, 2013).

Prior to the trial, five of the six accused publishers settled with the D.O.J. Hachette, HarperCollins, and Simon & Schuster were first, jointly settling on September 14th, 2012. These three publishers agreed to create a $69 million fund for refunds to customers. The largest of the five settlements was with Penguin for $75 million on May 25, 2013. Macmillan was the last to settle on August 16, 2013 for $26 million. Under the terms of these settlements, the publishers agreed to terminate their agency contracts with Apple (and all other e-book retailers), avoid setting retail prices in any future e-book agreements for two
years, and avoid sharing sensitive competitive information with fellow publishing companies for five years. While not issuing an outright ban of the agency e-book retail model, these terms provided by the D.O.J. forced the accused publishers to permit all retailers the right to discount the retail prices of e-books at their own discretion for two years, allowing Amazon to return to their favored $9.99 price point for new releases.

By the time of the trial on June 3, 2013, Apple was the only remaining defendant in the case. Apple defended their contracts with the accused publishers as necessary for the company’s entry into the e-book retail industry. Apple argued that the introduction of their iBookStore infused some much needed competition into the market, which had previously been dominated by Amazon, who with a 80-90 percent estimated market share had monopolistic control over e-book prices to the detriment of both the publishers and the consumers. In addition, Apple contended that their entry into the e-book industry inspired the growth of the e-book market from 400,000 cumulative downloaded e-books in 2010 to more than 1.7 million by the time of the trial in 2013. Accordingly, the D.O.J. was misguided in targeting a small competitive entrant with an aggressive antitrust lawsuit and not the monopolistic market-leader, Amazon.

New York District Judge Denise Cote delivered her Opinion to the Court on July 10, 2013, ruling that Apple's most-favored-nation clause enforcing the agency model on all e-book retailers defeated price competition (per se price fixing) and resulted in higher prices across the industry (price fixing in violation of the rule of reason). Cote made clear that although evidence that Apple’s price fixing had elevated prices was provided by the D.O.J., Apple’s guilt could be decided on a per se basis. Her verdict stated that Apple and the five publishers had agreed to “act collectively to force up Amazon's retail price” in violation of Section 1 of the Sherman Act (1890). As punishment, Apple had to terminate its contracts with the five accused publishers and not enter into any similar contracts with providers of music, movies, T.V., or apps for their iTunes Store for five years from the time of the judgment. Additionally, to increase e-book price transparency for customers, Apple was forced to insert into their iBookStore links to Amazon and Barnes & Noble's prices for the same e-book. Cote also found Apple liable to 33 states that joined the D.O.J. as
prosecutors in this suit, and has since set a preliminary date of May, 2014 for a secondary trial to determine Apple’s further damages and fines.

As a last twist, around this same time (July 1, 2013), Penguin Group Inc. and Random House L.L.C. merged to form Penguin Random House Inc. The Big Six Publishers were thus renamed the Big Five. Upon the completion of this merger, Penguin Random House had the largest market-share of both the electronic and print book publishing industries. As such, this merger has been referred to by some as the publishing industry’s response to the verdict of *U.S. v. Apple Inc. et. al. (2013)* and the increasing dominance of Amazon in the electronic and print book markets. Through this merger, the newly-formed Penguin Random House dramatically increased their leverage in negotiating higher e-book prices with Amazon to the benefit of the publishing industry at large (Ciabattari, 2013).

Apple appealed Judge Cote’s verdict on February 24, 2014. In the *Appellate Opening Brief* filed by Apple with the U.S. Court of Appeals for the Second Circuit of New York, representatives from Apple explained that their basis of appeal is on the grounds of misconstrued liability. They argue that although Apple had signed agency contracts with the five accused publishers, they were ignorant of the most-favored-nation clause enforcing the agency distribution model on the other retailers. The D.O.J. is expected to release their response in May, 2014. The appeals case is *U.S. v. Apple Inc., U.S. Court of Appeals for the Second Circuit (Manhattan)*.

**Section 3: Literature Review**

This case study analyzing the effect of the *U.S. v. Apple Inc. et. al. (2013)* verdict on e-book retail prices follows from a long history of relevant literature examining cartels and price collusion. A cartel is defined as a group of two or more independent sellers who explicitly agree to fix prices or output in a given market for the purpose of increasing their collective profits (Dick, 1998). Industrial-organization economic theory tells us that cartel collusion can allow firms to elevate prices above their competitive level. Generally, the participating firms agree to reduce output and/or fix higher prices to achieve closer to joint
profit maximization instead of the zero industry profit realized in a perfectly competitive market. On even a simple linear supply and demand graph, we can see how this reduction in industry output reduces the equilibrium quantity and increases the equilibrium price. This is an inefficient allocation of goods resulting in elevated industry profits, larger dead-weight loss, and reduced societal welfare, all to the detriment of the consumer (Pepall et. al., 2008). A supply and demand graph illustrating this point can be seen in Figure 1.

If the price-increasing effects of cartels was relatively small, their persistence might not warrant much concern. However, Froeb et. al. (1993) found that a price-rigging scheme involved in supplying frozen fish to the U.S. military elevated prices by 23-30 percent. Connor (2001) found that an international cartel of lysine manufacturers raised the market price of lysine by 17 percent, while Morse and Hyde (2000) found the effect to be twice as high at 34 percent (see below). In an extensive review of all cartel cases in the 20th century, Connor and Lande (2004) found that the median cartel price increase over all time-periods and across all cartel types was 22 percent. They estimated that this effect was 18 percent for domestic cartels and 32 percent for international cartels.

Prior to World War II, only the U.S. had any effectively-enforced antitrust laws, the Sherman Antitrust Act drafted in 1890, but there were few significant U.S. prosecutions until the mid-1940s (Berge, 1944; Stocking and Watkins, 1946). From 1943 to 1949, the U.S. Department of Justice (D.O.J.) convicted dozens of international cartels and won nearly all of their criminal cases (Wells, 2002). By the mid-1960s at least two dozen countries had antitrust laws and were serious in their efforts to enforce them (Edwards, 1967). Gallo et. al. (2000) have reviewed the enforcement of the Sherman Act by the D.O.J. from 1955 to 1997. By 1996, 70 countries, comprising about 78 percent of the world's output and 86 percent of the world's trade, had adopted competition laws (Palim, 1998).

Since 1995, the D.O.J. has had notable success in prosecuting international cartels due to several amendments to U.S. antitrust laws and improved investigative techniques (Connor, 2001; Baker, 2001). The Sherman Act's penalties, including corporate fines and participant prison sentences, have steadily increased by amendments in 1955, 1974, 1987, and 1990 (Connor, 2003). Since 1996, nearly 200 international cartels have been
uncovered and prosecuted by the D.O.J., the Competition Policy Directorate of the European Commission, and other antitrust authorities around the world (Connor and Helmers, 2006).

Beginning in 1995, the D.O.J.’s Antitrust Division made the prosecution of international cartels that victimize U.S. businesses and consumers one of its highest priorities, publicly announcing its plan to devote about 30 percent of all of its resources to criminal price-fixing prosecution (Connor, 2004). Similarly in Europe, Neelie Kroes who was appointed the European Union Competition Commissioner in 2004, publicly announced her “zero-tolerance” approach to prosecuting cartels. Even so, Evenett et. al. (2001) and Connor and Lande (2004) have found that cartel overcharges have remained so high and conspiracies so durable that current U.S. public and private monetary sanctions provide inadequate deterrence to their formation, meaning the benefits from realizing cartelized profits greatly exceeds the costs associated with losing an antitrust lawsuit.

The major role played by economic analysis in horizontal (within-industry) price-fixing cases is the calculation of the overcharge on buyers in markets affected by a cartel. The overcharge is the value of purchases of a cartelized product actually made, the “cartel price,” minus what the value of purchases would have been for the same volume of product absent the cartel, the “but-for price.” Accurate estimates of conspiracy-induced overcharges are primarily of importance in determining the recovery of civil damages and the calculation of U.S. government fines. Recently, there has been a lot of renewed interest in this area by economists (e.g. Harrington, 2004a; Harrington, 2004b; Levenstein and Suslow, 2004; Connor, 2004; Connor, 2001; Clarke and Evenett, 2003).

There has been much research on the construction of models estimating but-for prices in antitrust matters (e.g. Baker and Rubinfeld, 1999; Finkelstein and Levenbach, 1983; Rubinfeld and Steiner, 1983). Reduced-form pricing regressions are commonly used in the detection and prosecution of cartels (Nieberding, 2006; Paha, 2011). Specifically, the reduced-form model, dummy variable approach is possibly the most popular method used to isolate the impact of the anticompetitive conduct on the price
under study after controlling for influences unrelated to such conduct (Finkelstein and Levenbach, 1983; Nieberding, 2006). Authors that have implemented a reduced form model, dummy variable approach to provide estimates of cartel overcharge include Connor (2001), Clarke and Evenett (2003), Frank and Schliffke (2013), Froeb et. al. (1993), and Morse and Hyde (2000).

For an example of this methodology, consider Morse and Hyde’s (2000) study on collusion by a global cartel of lysine manufacturers. For data, this study had access to monthly lysine prices for the relevant firms from 1992 to 1995. Lysine is an amino acid frequently used as an additive in livestock feed. Included in the authors’ reduced form model were controls for lysine demand such as the number of hogs needed by U.S. slaughterhouses, red meat and poultry export demand, the price of complement and substitute goods, and the seasonality of lysine demand. Also included were controls for lysine supply such as the cost of a principal input in lysine production (dextrose) and other variable costs of manufacture and capital.

Morse and Hyde (2000) found that the coefficient on the dummy variable of interest in their regression, set to unity during the time of collusion, was 34 percent. This estimate suggests that lysine prices were 34 percent higher during the cartel’s existence ceteris paribus. Multiplying by the total quantity of lysine sold by the cartel from 1992 to 1995 provided an estimation of total cartel overcharge during the period of $162 million. This estimation of cartel overcharge could then be compared to the $70 million fine imposed on members of the cartel by the U.S. D.O.J. in settlement. Here, Morse and Hyde (2000) were able to definitively show that these cartel members were inadequately punished.

My research design relates most directly to the methodology laid out by Orley Ashenfelter (2013) in his Direct Testimony presented on behalf of the D.O.J. at the trial of U.S. v. Apple Inc. et. al. (2013). Ashenfelter’s model follows from the previous literature on estimations of cartel overcharge using a reduced form method, dummy variable approach (Nieberding, 2006; Rubinfeld and Steiner, 1983), but he extends this methodology to implement a difference-in-differences analysis.

Ashenfelter’s (2013) analysis estimated the change in e-book prices for the
conspiring publishers from the six-month period prior to the implementation of their agency model on April 1, 2010 (“pre-period”) to the six-month period following its implementation (“post-period”) with Random House acting as a control group from which to compare the accused publishers. Random House is a viable control group because they never entered into agency contracts with any of the e-book retailers and were never accused of violating Section 1 of the Sherman Act by the D.O.J. Ashenfelter estimated a 16.8 percent increase in the average e-book retail price across all retailers and accused publishers relative to Random House due to the adoption of the accused publishers’ agency contracts.

For Ashenfelter (2013), the collusion dummy variable is set to zero during the time of competition, prior to the accused publishers’ adoption of their agency contracts, and to unity during the time of collusion, after the agency contracts were signed. The coefficient on this dummy variable shows the effect of the collusive period on e-book prices ceteris paribus, with a null hypothesis of no effect (zero slope) and an alternative hypothesis of a non-zero effect on prices (specifically greater than zero). A coefficient significantly different from zero in the positive direction on this collusion dummy variable would show that relative to the Random House, the average price for e-books published by the accused publishers was higher during the time of collusion, after the agency contracts were signed, compared to the time of competition.

My analysis mirrors that of Ashenfelter’s (2013) to complete the story. Whereas Ashenfelter was attempting to estimate the increase in e-book retail prices due to the adoption of the accused publishers’ agency contracts, I attempt to estimate the subsequent decrease in e-book retail prices due to the D.O.J.’s intervention. Ashenfelter and I, therefore, apply analogous methodologies to different datasets and different time-periods, studying the same industry and surrounding the same trial.

In this study, the collusion dummy variable is set to unity when competition was present in the e-book retail market, after the accused publishers settled with the D.O.J., and to zero when collusion was present in the market, when the accused publishers had agency contracts in effect with their retailers. The coefficient on this dummy variable shows the effect of the collusive period on e-book prices ceteris paribus, with a null hypothesis of
no effect (zero slope) and an alternative hypothesis of a non-zero effect on prices (specifically less than zero). Here, a coefficient significantly different from zero in the negative direction on this collusion dummy variable would show that average e-book prices for e-books published by the accused publishers were lower during the time of competition, after the settlements of the accused publishers, compared to the time of collusion, when their agency contracts were still in effect.

For that reason, this study adds to the previous literature by applying an estimation of cartel overcharge methodology to the settlement of the accused publishers in *U.S. v. Apple Inc. et. al. (2013)*. Providing evidence of a significant overcharge during the agency-model period would be important, because it would add credence to the verdict of the District Court in ruling Apple Inc. guilty of price fixing. This would serve as counter-evidence to Apple’s continued plea of innocence, who on February 25, 2014 appealed this verdict.

This study also adds to the previous literature on difference-in-differences analysis. Classical difference-in-differences analyses, such as in the model implemented by Ashenfelter (2013), use a “treatment group” and a “control group”, a “pre-period” and a “post-period” to construct the four necessary categories necessary for the D.i.D. comparison (Card and Krueger, 1994). This study constructs a more complex, multi-dimensional difference-in-differences analysis. Here, the sample is split into five separate treatment groups, corresponding to each of the five accused publishers, with Random House serving as the control group. Similarly, the sample is split into four separate time-periods, corresponding to each of the four distinct time-periods surrounding the three publisher settlement dates. This form of difference-in-differences analysis allows for me to manually compare the change in e-book prices for any one of the five accused publishers relative to Random House from any one of the four time-periods to any of the others using Stata’s post-regression linear-combination (“lincom”) feature. In this way, this analysis adds to the literature on the application of extended difference-in-differences models.
Section 4: Data  
Section 4a: Primary Dataset

For his data, Ashenfelter (2013) used subpoenaed historical e-book price and sales data supplied by each of the major e-book retailers specifically for the trial. Unfortunately, the e-book industry is notoriously tight-lipped when it comes to sales statistics, and I do not have access to the same dataset as Ashenfelter for the more-recent time-period in which I am interested.

Instead, the primary dataset used in this analysis was received from iobyte Solutions, Inc. (http://www.iobyte.com/), an I.T. consulting company based in Thornwood, New York. iobyte Solutions provides a service called “e-book Market View” for its clients, primarily to help publishers analyze “competitive pricing behavior, metadata compliance, author discovery, and merchandising analytics.” As a back-end to this service, iobyte Solutions has compiled a large database of historical e-book retail prices by scraping all of the major e-book retailers daily. The data was provided by Dan Lubart, a Managing Partner at iobyte Solutions, on January 12, 2014.

This dataset includes 91 weeks of e-book retail pricing data ranging from April 8, 2012 to December 29, 2013 for a sample of 500 ISBNs for the three largest e-book retailers: Amazon, Apple, and Barnes & Noble and all six of the Big Six publishers. For each week, I have data on these same 500 e-books for each of these three retailers, giving me theoretically 1,500 unique observations (ISBN-retailer pairs) each week, and a total of 136,500 unique observations (ISBN-retailer-week triples). In my actual dataset, I have significantly fewer observations than this, only 26,821 unique observations (ISBN-retailer-week triples), because all 500 e-books are not present every week and I do not have data for e-books sold at Apple for every week. Each row of this dataset corresponds to the retail price for an ISBN being sold at one of the aforementioned retailers in a given week. For clarity purposes, I therefore refer to a single price observation (i.e. a single row of my dataset) as the price of e-book $i$, with retailer $j$, in week $t$.

For each ISBN-retailer-week observation in this iobyte Solutions dataset, I have the following variables: e-book price in U.S.D. ($PRICE_{ijt}$), e-book publisher ($PUBLISHER_i$),
and e-book sales ranking ($RANK_{ijt}$). An e-book's sales ranking is the ordinal sales-rank of e-book $i$, with retailer $j$, during week $t$. For example, ISBN 9781612130293 might be the 15th best-selling e-book for Amazon in week four.

**Section 4b: Secondary Dataset**

To supplement the e-book retail pricing data from iobyte Solutions Inc., I queried the Google Books A.P.I. for additional e-book level information for the 500 ISBNs in my iobyte Solutions dataset. Google Books is a service offered by Google, Inc. allowing users to search and preview millions of books from libraries and publishers all over the world (http://books.google.com/). As part of the service, Google provides the Google Books A.P.I. which allows developers restricted access to Google's database of books serving as the back-end to all Google Books features.

Using my personal Google/GMail credentials, I was able to request an A.P.I. key from Google allowing me restricted access to the Google Books A.P.I. (less than 5,000 requests per day). For each unique ISBN in the iobyte Solutions dataset, I used a JavaScript script to query the Google Books A.P.I. with a given ISBN to retrieve additional information about this e-book. The data was retrieved from Google on February 18, 19, and 20, 2014. This process could be repeated by anyone with a Google Account by referencing the Google Books A.P.I. Documentation page (https://developers.google.com/books/).

For each ISBN in this dataset, I have the following variables: e-book genre ($GENRE_i$), e-book page count in its corresponding print version ($PAGECOUNT_i$), e-book publication date ($PUBLICATIONDATE_i$), e-book average rating among Google Books users ranging from 1.0 to 5.0 ($RATING_i$), and the number of ratings submitted by Google Books users in determining the e-book's Google Books average rating ($RATINGSCOUNT_i$).

As an e-book's ISBN number is a unique identifier for that e-book, this supplementary Google Books dataset could then be merged with the dataset from iobyte Solutions Inc. using each e-book’s ISBN as the primary key to form the final dataset used in...
this analysis.

Section 5: Theory

In this section I derive the reduced form model, dummy variable approach (Nieberding, 2006), difference-in-differences analysis (Ashenfelter, 2013) as used in this study. Consider that prices are determined through the simultaneous interaction of demand and supply factors in a market and that demand and supply can be represented by the following two example estimating equations (Nieberding, 2006).

\[ Q_t^D = \alpha_0 + \alpha_1 P_t + \alpha_2 X_t^D + \epsilon_t \]  \hspace{1cm} (1)
\[ Q_t^S = \delta_0 + \delta_1 P_t + \delta_2 X_t^S + \epsilon_t \]  \hspace{1cm} (2)

Here \( Q_t^D \) and \( Q_t^S \) represent the quantity demanded and supplied respectively at time \( t \), \( P_t \) is the market price at time \( t \), and \( X_t^D \) and \( X_t^S \) represent linear exogenous demand and supply influences respectively. For example, \( X_t^D \) could represent average income in the geographic area of interest at time \( t \), and \( X_t^S \) could represent an essential input cost for production at time \( t \). The \( \epsilon_t \)'s are independent, identically distributed error terms with mean of zero.

The reduced-form model uses the equilibrium condition that market demand equals market supply (\( Q_t^D = Q_t^S \)) to solve the above two equations for equilibrium price (\( P_t \)).

\[ P_t = \beta_0 + \beta_1 X_t^D + \beta_2 X_t^S + \epsilon_t \]  \hspace{1cm} (3)

This yields a reduced-form equation that directly relates equilibrium price to the above exogenous demand and supply factors. Note that the coefficients on each of the independent variables in Equation 3 are also intrinsically related to the corresponding coefficients in Equations 1 and 2.
\[ \beta_0 = \frac{\delta_0 - \alpha_0}{\alpha_1 - \delta_1} \]  
\[ \beta_1 = \frac{\alpha_2}{\delta_1 - \alpha_1} \]  
\[ \beta_2 = \frac{\delta_2}{\alpha_1 - \delta_1} \] 

(3a) (3b) (3c)

The dummy variable approach requires data points for all time-periods in which there is data, both conspiratorial and non-conspiratorial (Rubinfeld and Steiner, 1983). The model includes one additional dummy variable, \( C \), equal to unity when the cartel was in effect and to zero during the time of competition in the market.

\[ P_t = \beta_0 + \beta_1 X_t^D + \beta_2 X_t^S + \beta_3 C + \varepsilon_t \]  

(4)

The coefficient on the cartel dummy variable, \( \beta_3 \), then indicates whether the average price during the cartel's existence was significantly different, and in particular greater than the average price during the competitive period (Rubinfeld and Steiner, 1983). This gives the null and alternative hypotheses for the model.

\[ H_0: \beta_3 = 0 \quad H_1: \beta_3 \neq 0 \]  

(4a)

Ashenfelter (2013) extended the reduced form method, dummy variable approach by implementing a difference-in-differences analysis with Random House acting as a control group from which to estimate the change in prices for the publishers accused of collusion. To model this, consider two firms in an industry, firm A and firm B, where firm A (for “accused”) was accused of collusion, while firm B was not. If we have data for both firms, we can implement a difference-in-differences analysis by including an “accused” dummy variable, \( A \), equal to unity if the price observation is for a good sold by firm A, and equal to zero otherwise (in this case, if the good was sold by firm B).

\[ P_t = \beta_0 + \beta_1 X_t^D + \beta_2 X_t^S + \beta_3 C + \beta_4 A + \beta_5 C*A + \varepsilon_t \]  

(5)
$\beta_5$, the coefficient on the interaction of the collusion dummy variable, $C$, with the accused dummy variable, $A$, serves as the difference-in-differences estimator for the change in the average price of goods sold by firm $A$ from the collusive time-period to the competitive time-period relative to the change in the average price of goods sold by firm $B$ over the same time-periods.

The fact that $\beta_5$ represents the difference-in-differences estimator for firm $A$ in Equation 5 can be shown by looking at which coefficients affect firm $A$’s and firm $B$’s prices in each time-period. The price of goods sold by firm $A$ ($A = 1$) in the collusive time-period ($C = 1$) is determined by Equation 5a below. The price of goods sold by firm $A$ ($A = 1$) in the competitive time-period ($C = 0$) is determined by Equation 5b. Similarly, the price of goods sold by firm $B$ ($A = 0$) in the collusive time-period ($C = 1$) is determined by Equation 5c, and the price of goods sold by firm $B$ ($A = 0$) in the competitive time-period ($C = 0$) is determined by Equation 5d.

\[
\begin{align*}
P_t^A &= \beta_0 + \beta_1 X_t^D + \beta_2 X_t^S + \beta_3 + \beta_4 + \beta_5 + \epsilon_t \quad (5a) \\
P_t^A &= \beta_0 + \beta_1 X_t^D + \beta_2 X_t^S + \beta_4 + \epsilon_t \quad (5b) \\
P_t^B &= \beta_0 + \beta_1 X_t^D + \beta_2 X_t^S + \beta_3 + \epsilon_t \quad (5c) \\
P_t^B &= \beta_0 + \beta_1 X_t^D + \beta_2 X_t^S + \epsilon_t \quad (5d)
\end{align*}
\]

The “first-difference” in this difference-in-difference analysis looks at the change in prices for firm $A$ and firm $B$ from the collusive time-period to the competitive time-period by subtracting Equation 5b from Equation 5a for firm $A$ and Equation 5d from Equation 5c for firm $B$. Equation 5e below shows this first-difference for firm $A$, and Equation 5f shows this first-difference for firm $B$.

\[
\begin{align*}
\Delta P_t^A &= \beta_3 + \beta_5 \\
\Delta P_t^B &= \beta_3
\end{align*}
\]

The “second-difference” or “difference-in-differences” is then calculated as the
change in prices for firm A from the collusive time-period to the competitive time-period (Equation 5e) minus the change in prices for firm B from the collusive time-period to the competitive time-period (Equation 5f). This difference-in-differences is determined by Equation 5g below.

\[ \Delta P_t^A - \Delta P_t^B = \beta_5 \]  

(5g)

Here, it should be clear that \( \beta_5 \) estimates the change in the average price of goods sold by firm A from the collusive time-period to the competitive time-period relative to the change in the average price of goods sold by firm B over the same time-periods. The null and alternative hypotheses for this difference-in-differences model (Equation 5) is:

\[ H_0: \beta_5 = 0 \quad \text{H}_1: \beta_5 \neq 0 \]  

(5h)

Ashenfelter’s model further extends this standard difference-in-differences example, however, because in the case of U.S. v. Apple Inc. et. al. (2013), there were five accused firms with one firm (Random House) serving as the control group. The five accused publishers were Hachette, HarperCollins, Penguin Group, Macmillan, and Simon & Schuster. The inclusion of Random House as the baseline group controls for many exogenous factors affecting price across the entire e-book industry such as changes in U.S. median income, changes in iPad sales over time, or changes in e-book tax rates.

If we have data for all of the firms and the accused firms all share the same collusion dummy variable (meaning they transitioned from collusion to competition at the same time), the model can easily be extended to estimate a difference-in-differences estimator for each of \( N \) accused firms. For the simple case where \( N = 2 \), consider two accused firms, \( A_1 \) and \( A_2 \), both being compared to one control firm, \( B \). The new model will look like the following:
\[ P_t = \beta_0 + \beta_1 X_t^D + \beta_2 X_t^S + \beta_3 C + \beta_4 A_1 + \beta_5 C*A_1 + \beta_6 A_2 + \beta_7 C*A_2 + \epsilon_t \]  \tag{6}

Following the same difference-in-differences process outlined in Equations 5a-5g, it can be shown that \( \beta_5 \) is the difference-in-differences estimator for firm \( A_1 \) and \( \beta_7 \) is the difference-in-differences estimator for firm \( A_2 \). That is, \( \beta_5 \) estimates the change in the average price of goods sold by firm \( A_1 \) from the collusive time-period to the competitive time-period minus the change in the average price of goods sold by firm \( B \) from the collusive time-period to the competitive time-period. \( \beta_7 \) estimates the same for firm \( A_2 \) relative to firm \( B \).

This model can be further extended to any number of accused firms, \( N \).

\[ P_t = \beta_0 + \beta_1 X_t^D + \beta_2 X_t^S + \beta_3 C + \beta_4 A_1 + \beta_5 C*A_1 + \beta_6 A_2 + \beta_7 C*A_2 + \ldots + \beta_N A_N + \beta_{N+1} C*A_N + \epsilon_t \]  \tag{7}

In Equation 7, we would have \( N \) null and alternative hypotheses, one for each of the interaction terms which serve as the difference-in-differences estimators for each of the \( N \) accused firms. The null and alternative hypotheses for Equation 7 would be the following:

\[ H_0: \beta_5 = 0 \quad H_1: \beta_5 \neq 0 \]  \tag{7a}
\[ H_0: \beta_7 = 0 \quad H_1: \beta_7 \neq 0 \]  \tag{7b}
\[ \ldots \]  
\[ H_0: \beta_{N+1} = 0 \quad H_1: \beta_{N+1} \neq 0 \]  \tag{7N}

Ashenfelter (2013) uses a model where \( N = 5 \), to calculate a difference-in-differences estimator for each of the five accused publishers. He can include one collusion dummy variable, \( C \), for all five of the accused publishers, because all five of the accused publishers adopted agency contracts on the same date (April 1, 2010).
In my analysis, there are multiple publisher-settlement dates, so I extend Ashenfelter’s (2013) model to incorporate multiple collusion dummy variables. This is my point of departure from the previous literature and Nieberding’s (2006) and Ashenfelter’s (2013) theoretical background. Multiple collusion dummy variables in an estimation of cartel overcharge model using a reduced form method, dummy variable approach with a difference-in-differences analysis adds complexity to the model, but it allows for me to look into the change in the average e-book retail price for any of the accused publishers between any two time-period pairs. This helps us better understand the movement of e-book retail prices over the time-period of interest and is novel to the literature.

Section 6: Methods

In my analysis, I run an O.L.S. regression with the natural log of the price in U.S.D. of e-book $i$, with retailer $j$, at week $t$ as the dependent variable to estimate the change in e-book retail prices for the accused publishers (Hachette, HarperCollins, MacMillan, Penguin Group, and Simon & Schuster) as they transitioned from a collusive period, defined as the time when their agency contracts were still in effect with their retailers, to a competitive period, defined as the time after each publisher settled with the U.S. Department of Justice. Similarly to Ashenfelter (2013), I use Random House as a control group for they never entered into agency contracts with any of the retailers and were never accused of violating Section 1 of the Sherman Act by the D.O.J.

The date of the verdict of U.S. v. Apple Inc. et. al. (2013) could be seen as the date at which all five of the accused publishers transitioned from their collusive period to the competitive period, but testimony at trial has suggested that the publisher settlement dates better represent the actual point in time where e-book retail prices changed, because as part of the terms of their settlement with the D.O.J., each of the accused publishers were required to re-allow retail-discounting on all of their e-books (Cote, 2013). This effectively negated their agency contracts and returned competition on price to the e-book retailer level. This complicates the model developed in Section 5 because different publishers settled with the D.O.J. on different dates. I therefore extend Ashenfelter’s (2013) single
collusion dummy variable model to include multiple collusion dummy variables.

Hypothetically, we could construct a model with $N$ accused firms and $N$ collusion dummy variables, but in the case of *U.S. v. Apple Inc. et. al.* (2013), there were five accused publishers with three distinct settlement dates among them, because Hachette, HarperCollins, and Simon & Schuster jointly settled with the D.O.J. on the same date. This means that the model should only include three collusion dummy variables, one for each of the three settlement dates. In each of their settlement agreements, the accused publishers were granted an approximately seven-day grace period by the D.O.J. to terminate their agency contracts and to re-allow retailer discounting. For this reason, I use the end of this grace period as the date at which the accused publishers switched from collusion to competition. Hachette (HA), HarperCollins (HC) and Simon & Schuster (SS) jointly settled on September 14, 2012 ($C_{HHS}$). Penguin Group (PG) settled on May 25, 2013 ($C_{PG}$), and Macmillan (MM) settled on August 16, 2013 ($C_{MM}$).

$N$ collusion dummy variables split the analysis into $N + 1$ distinct time-periods. For example, consider that one collusion dummy variable results in two time-periods: a collusive time-period and a competitive time-period. Similarly, three collusion variables split the analysis into four time-periods: $W$, $X$, $Y$, and $Z$. The first time-period, $W$, corresponds to all observations ranging from April 8, 2012 to September 13, 2012, from the beginning of my sample to just prior to the first settlement date, when all of the accused publishers still had agency contracts in effect. The second time-period, $X$, corresponds to all observations ranging from September 14, 2012 to May 24, 2013, from the settlement date of Hachette, HarperCollins, and Simon & Schuster to just prior to the settlement date of Penguin Group, when only Penguin Group and Macmillan still had agency contracts in effect. The third time-period, $Y$, corresponds to all observations ranging from May 25, 2013 to August 15, 2013, from the settlement date of Penguin Group to just prior to the settlement date of Macmillan, when only Macmillan still had agency contracts in effect. The fourth and final time-period, $Z$, corresponds to all observations ranging from August 16, 2013 to December 29, 2013, from the settlement date of Macmillan to the end of my sample, when all publishers had settled with the D.O.J. and nullified their agency contracts.
These time-periods are graphed in Figure 3.

Partitioning my dataset into these four time-periods, \( W, X, Y, \) and \( Z \), allows for me to not only study the effect of each publisher's settlement date on their e-book prices, but the cross-effect of publisher \( B \)'s settlement date on publisher \( A \)'s e-book prices. In this way, I can control for the effect of all of the other publishers' settlement dates on the e-book prices for any given publisher. If I were to instead split the dataset into only two time-periods for each publisher along their settlement date, I would be making the assumption that the effect of all of the other publishers' settlement dates on this publisher's e-book prices was zero.

It may be that all of the publishers really abandoned their agency contracts after the first settlement date (Hachette, HarperCollins, and Simon & Schuster). Alternatively, it may be that publisher \( B \)'s settlement date had no effect on publisher \( A \)'s e-book prices. For example, my model can estimate the effect of HarperCollins’ settlement with the D.O.J. on Penguin Group's e-book prices, by looking at the change in prices for e-books published by Penguin Group from time-period \( W \) to time-period \( X \).

To test these hypotheses, there are three required steps in calculating the difference-in-differences estimator for each of the five accused publishers: \( HA, HC, MM, PG, \) and \( SS \), for the change in price between any of the four time-periods: \( W, X, Y, \) and \( Z \). First, I show which coefficients will be included in determining the price of e-books published by each publisher in each time-period in Table 1. Depending on the publisher of the e-book and the date of the price observation, many terms drop out of the model so that only the factors that directly relate to publisher \( A \)'s e-book prices are used for its estimation.

Second, I take the first-difference by calculating the difference in each publisher's coefficients between each of the two time-period combinations (\( X - W, Y - W, Z - W, Y - X, Z - X, Z - Y \)) in Table 2. For example column one of Table 2, \( X - W \), shows which coefficients determine the change in e-books prices for publisher \( A \) from time-period \( W \) to time-period \( X \).

The third and final step is to take the second-difference of the change in e-book
prices for each of the accused publishers minus the change in e-book prices for Random House (the control group) over the same time-periods. This is presented in Table 3. Each entry shows which coefficients determine the difference in prices for publisher A in time-period \( N \) minus the prices for publisher A in time-period \( N - 1 \) and the the prices for Random House in time-period \( N \) minus the prices for Random House in time-period \( N - 1 \). Each of these entries therefore represents the D.i.D. estimator for an accused publisher across two time-periods. For example row one, column one of Table 3 shows the difference-in-differences estimator for Hachette relative to Random House from time-period \( W \) to time-period \( X \). This D.i.D. estimator is captured by the coefficients: \( \beta_1 - \beta_{14} \). As will be explained further in Section 7, my null hypothesis for this D.i.D. estimator is that \( \beta_1 - \beta_{14} = 0 \), with the alternative hypothesis that \( \beta_1 - \beta_{14} \neq 0 \). Similarly, for each of the other respective entries.

It is important to note that in retrieving the above difference-in-differences estimators only the actual coefficient estimates (\( \beta_N \)'s) are calculated in the following regressions presented in Sections 6a, 6b, and 6c. The first-difference and second-difference for each publisher and time-period pair are calculated manually by using Stata's prepackaged, post-regression, linear-combination ("lincom") feature.

I construct three models. The first includes just the essential variables needed for the difference-in-differences analysis measuring the change in e-books prices for the accused publishers from the pre-period (prior to settlement with the D.O.J.) to the post-period (after settlement) relative to the change in e-book prices for Random House over the same periods. The second model uses these same variables but adds ISBN-level and retailer-level fixed effects. The third model has the same structure as Model II except implements several ISBN-level controlling variables instead of ISBN-level fixed effects. Each of these models is explained in Section 6a, 6b, and 6c respectively. My hypotheses for these three models are then presented in Section 7.
Section 6a: Model I

\[ \text{LPrice}_{ijt} = \beta_0 + \beta_1 \text{HA} \cdot \text{C}_{\text{HHS}} + \beta_2 \text{HA} + \beta_3 \text{C}_{\text{HHS}} + \beta_4 \text{HC} \cdot \text{C}_{\text{HHS}} + \beta_5 \text{HC} + \beta_6 \text{MM} \cdot \text{C}_{\text{MM}} + \beta_7 \text{MM} + \beta_8 \text{C}_{\text{MM}} + \beta_9 \text{PG} \cdot \text{C}_{\text{PG}} + \beta_{10} \text{PG} + \beta_{11} \text{C}_{\text{PG}} + \beta_{12} \text{SS} \cdot \text{C}_{\text{HHS}} + \beta_{13} \text{SS} + \beta_{14} \text{RH} \cdot \text{C}_{\text{HHS}} + \beta_{15} \text{RH} \cdot \text{C}_{\text{MM}} + \beta_{16} \text{RH} \cdot \text{C}_{\text{PG}} + \varepsilon_{ijt} \]

PRICE is a numerical variable measuring the price in U.S.D. of e-book \( i \), with retailer \( j \), in week \( t \). The average price of e-books in my sample is $9.29, but these prices range from $0.00 (free) to $44.99. As can be seen in the Data Appendix, the distribution of PRICE is right skewed as 98.42 percent of e-books in my sample are priced in the $0.00 to $14.99 range, while a couple hundred outliers (1.58 percent of all e-books) are priced higher. To correct this skewed distribution, I use the natural log of PRICE. Since some e-books in my sample are free to purchase (i.e. PRICE = $0.00), I add one to PRICE before taking its natural log, because the natural log of zero is undefined, while the natural log of one is zero. Like Ashenfelter (2013), I therefore use LPRICE as the dependent variable in my model, calculated as one plus the natural log of PRICE. This transformation should also help to mitigate any heteroscedasticity issues in the dependent variable.

To control for influences specific to each publisher, I include publisher fixed effects. This means I include five dummy variables, one for each of the accused publishers: Hachette (HA), HarperCollins (HC), Macmillan (MM), Penguin Group (PG), and Simon & Schuster (SS), with Random House serving as the omitted state.

To measure the effect of each publisher's settlement date on the prices of e-books published by that publisher, I include one collusion dummy variable for each of the three settlement dates. These collusion dummy variables are equal to unity if the e-book price observation date is within the competitive period for the publisher of that e-book, after that publisher settled with the D.O.J., and zero if it is within their collusive period, when agency contracts were still enforced by the publisher.

Like the publisher fixed effects above, Random House again serves as the omitted
state for these collusion variables, so does not have an associated collusion dummy variable. This is intuitive because Random House was never accused of collusion in violation of Section 1 of the Sherman Act, so never settled with the D.O.J. Since Hachette, HarperCollins, and Simon & Schuster jointly settled with the D.O.J., these three publishers share a collusion dummy variable ($C_{HHS}$).

Model I includes the interaction between the five accused publisher dummy variables with their corresponding collusion dummy variable, as well as the separate, non-interacted publisher and collusion dummy variables. The inclusion of an interaction term with each of the interacting dummy variables separately in the regression is common in D.i.D. analysis for interpreting the D.i.D. coefficient (Card and Krueger, 1994). I also include Random House interacted with each of the three collusion dummy variables to measure the change in the average price of e-books published by Random House over the time-periods before and after each of the accused publishers' settlements so that Random House can serve as the baseline in my difference-in-differences approach.

As was explained previously, the calculation of the difference-in-differences estimators requires three steps. First, we calculate the factors (or coefficients) involved in determining e-book $i$'s price for each of the six publishers and in each of the four time-periods. This is presented in Table 1. Next, we subtract the average e-book prices of publisher $A$ in the post-period from the average e-book prices of publisher $A$ in the pre-period. Note, we also subtract the average e-book prices of Random House in the same post-period from the average e-book prices of Random House in the same pre-period. This is presented in Table 2. Lastly, we subtract both of these “first differences,” the change in e-book prices for publisher $A$ from the change in e-book prices for Random House to calculate the “second differences”, also referred to as the “difference-in-differences estimator”: the change in e-book prices for publisher $A$ from their corresponding pre-period to their corresponding post-period relative to the change in e-book prices for Random House, the control group, over the same time-periods. The coefficients determining the difference-in-differences estimators for each of the accused publishers are presented in Table 3.
Section 6b: Model II

\[ \ln(\text{Price}_{ijt}) = \beta_0 + \beta_1 \text{HA} \times \text{C}_{HHS} + \beta_2 \text{HA} + \beta_3 \text{C}_{HHS} + \beta_4 \text{HC} \times \text{C}_{HHS} + \beta_5 \text{HC} + \beta_6 \text{MM} \times \text{C}_{MM} + \beta_7 \text{MM} + \beta_8 \text{C}_{MM} + \beta_9 \text{PG} \times \text{C}_{PG} + \beta_{10} \text{PG} + \beta_{11} \text{C}_{PG} + \beta_{12} \text{SS} \times \text{C}_{HHS} + \beta_{13} \text{SS} + \beta_{14} \text{RH} \times \text{C}_{HHS} + \beta_{15} \text{RH} \times \text{C}_{MM} + \beta_{16} \text{RH} \times \text{C}_{PG} + \beta_{17} \text{APPLE} + \beta_{18} \text{BNOBLE} + \sum_{k=1}^{499} \beta_{18+k} \text{BOOK}_k + \epsilon_{ijt} \]

In addition to the primary variables necessary for the difference-in-differences analysis included in Model I, the nature of my dataset allows for me to include ISBN-level and retailer-level fixed effects analogous to those included in Ashenfelter’s (2013) study. I can include ISBN-level fixed effects, because I am tracking the prices of the same 500 e-books over time. I can include retailer-level fixed effects, because each e-book price observation over time is associated with both a specific ISBN and a specific retailer, either Amazon, Apple, or Barnes & Noble.

To implement retailer-level fixed effects, \text{APPLE} and \text{BNOBLE} are constructed as retailer dummy variables equal to unity if the e-book is being sold by Apple or Barnes & Noble respectively. Amazon serves as the omitted state. To implement ISBN-level fixed effects, \text{BOOK}_k is constructed as a vector of the 500 ISBNs included in the dataset with the lowest ISBN chosen as the omitted state.

Note that the inclusion of ISBN-level fixed effects and retailer-level fixed effects does not change the primary difference-in-differences analysis conducted for Model I, so Tables 1, 2, and 3 still apply to Model II in determining each publisher’s D.i.D. estimator for any two time-period pairs.
Section 6c: Model III

\[ \text{LPrice}_{ijt} = \beta_0 + \beta_1 HA*C_{HHS} + \beta_2 HA + \beta_3 C_{HHS} + \beta_4 HC*C_{HHS} + \beta_5 HC + \beta_6 MM*C_{MM} + \beta_7 MM + \beta_8 C_{MM} + \beta_9 PG*C_{PG} + \beta_{10} PG + \beta_{11} C_{PG} + \beta_{12} SS*C_{HHS} + \beta_{13} SS + \beta_{14} RH*C_{HHS} + \beta_{15} RH*C_{MM} + \beta_{16} RH*C_{PG} + \beta_{17} APPLE + \beta_{18} BNOBLE + \beta_{19} FICTION + \beta_{20} JUVEFICTION + \beta_{21} LAGE + \beta_{22} LRANK + \beta_{23} LPAGECOUNT + \beta_{24} RATING + \beta_{25} LRATINGSCOUNT + \beta_{26} MERGER + \epsilon_{ijt} \]

Alternatively to using ISBN-level fixed effects as in Model II, the data from the Google Books A.P.I. allows for the inclusion of several ISBN-level controls in Model III. In this model, I still include the publisher fixed effects from Model I and the retailer fixed effects from Model II, but I now include ISBN-level controls instead of ISBN-level fixed effects. Note that similarly to Model II, the addition of these controls does not change the primary difference-in-differences analysis conducted for Model I, so Tables 1, 2 and 3 still apply to Model III. Each of the ISBN-level controls unique to Model III is described in turn below.

**GENRE** is a categorical variable detailing the genre of e-book \( i \). To ensure that each genre-category is large enough (meaning it has a significant number of ISBN observations), I include only three genres: fiction, juvenile fiction, and other. In my model, I therefore convert **GENRE** into a vector of dummy variables. I include a dummy variable for fiction, **FICTION**, and a dummy variable for juvenile fiction, **JUVEFICTION**, with the genre **OTHER** serving as the omitted state.

**AGE** is the age of e-book \( i \) calculated as the difference between the date of the price observation and the date of publication for the corresponding print version of e-book \( i \). I take the natural log of **AGE** because as can be seen in the Data Appendix, **AGE** is right skewed (most e-books in my sample are relatively young, but several are extremely old). Similarly to **LPRICE**, I add one to **AGE** before taking the natural log because the minimum value of **AGE** is zero, meaning a price observation for an e-book occurred within the same week of its publication, and the natural log of zero is undefined.
RANK is the ordinal sales rank of e-book \( i \), with retailer \( j \), in week \( t \). For example, if ISBN 9781612130293 has a sales rank of 15 with Amazon in week four, then that means that this e-book was the 15th best-selling e-book for Amazon on that date. Similarly to AGE, RANK is right skewed so I include the natural log, LRANK, in my model.

PAGECOUNT is the number of pages in the corresponding print version of e-book \( i \). Similarly to RANK, this variable is right skewed so I include the natural log, LPAGECOUNT, in my model.

RATING is the average rating for e-book \( i \) as voted by Google Books users. If you have a Google Books account, you can rate any of the books listed in the Google Books database on a 1.0 to 5.0 scale. RATING is the mean rating across all Google Books users on the date that this dataset was retrieved.

RATINGSCOUNT is the total number of Google Books users who rated e-book \( i \), i.e. the number of votes used in calculating e-book \( i \)’s RATING. This is important, because it makes a big difference if only a handful of votes were used in calculating an e-book’s RATING versus a couple hundred. Like AGE, RATINGSCOUNT is right skewed and has a minimum of zero, so I include the natural log of one plus RATINGSCOUNT, LRATINGSCOUNT, in my model.

MERGER is a binary variable equal to unity after July 1, 2013, the date of the Penguin Group – Random House merger, and to zero before. This variable also roughly corresponds to the date of the trial, which began on June 3, 2013 and ended on July 10, 2013. This dummy variable should control for changes in e-book prices associated with these events.

Section 7: Hypotheses

The purpose of my three models is to estimate the change in e-book prices for accused publisher \( A \) from the pre-period before their settlement date to the post-period after their settlement date relative to the change in e-book prices for Random House across the same pre-period and post-periods. Generally, I predict that relative to Random House, the e-book retail prices for each of the accused publishers fell as they moved from
their respective pre-periods into their post-periods, after they settled with the D.O.J. and disbanded their agency contracts.

Significant, negative results would provide evidence in support of my hypothesis: Apple’s agency contracts with the accused publishers elevated prices across the e-book retail market and after each publisher’s settlement with the D.O.J. these prices came down. Further, the average price-decrease across publishers can be compared to Ashenfelter’s (2013) estimate of a 16.8 percent increase in e-book retail prices at the onset of these agency contracts. If my average D.i.D. price-decrease estimate is insignificantly different from Ashenfelter’s, it provides further evidence that post-agency e-book retail prices have returned to their competitive, pre-agency levels. These results would provide strong evidence for the correctness of the District Court’s verdict, the success of the D.O.J.’s lawsuit, and the misguided intent of Apple to appeal the case.

I test the null hypothesis that the change in e-book prices for publisher A was equal to the change in e-book prices for Random House over the same time-period. My alternative hypothesis is that the change in e-book prices for publisher A from the collusive time-period to the competitive time-period as defined by their settlement date is not equal to the change in e-book prices for Random House over the same time-periods. I test for strict inequality in my alternative hypothesis, because it is harder to prove, but I am really interested in the one-sided inequality that this difference-in-differences comparison is negative, meaning that e-book retailer prices decreased for each of the accused publishers from their pre to post-periods relative to Random House.

I have four distinct time-periods for each of the accused publishers instead of just two (a fixed collusive and competitive period), so I have choices in how I define each publisher’s pre and post-period. With my models, I can predict the difference-in-differences estimator for the change in e-book prices for each publisher for any two time-periods: \( X - W, Y - W, Z - W, Y - X, Z - X, \) and \( Z - Y. \) I construct three hypotheses corresponding to the different combinations of the time-period pairs I can use as pre and post-periods for each of the accused publishers below.


Section 7a: Hypothesis I

For each of the five accused publishers, I predict that the prices of e-books published by that publisher decreased relative to Random House from the time-period immediately before their settlement date to the time-period immediately after. This means that I predict that the prices of e-books published by Hachette, HarperCollins, and Simon & Schuster decreased from time-period $W$ to time-period $X$, the prices of e-books published by Penguin Group decreased from time-period $X$ to time-period $Y$, and the prices of e-books published by Macmillan decreased from time-period $Y$ to time-period $Z$.

This hypothesis isolates the change in firm $A$’s e-book retail prices due to firm $A$’s settlement with the D.O.J. ceteris paribus by only looking at the two time-periods in which their collusion dummy variable switches. All other publishers’ collusion dummy variables are held constant over this period, because they either have already settled or have yet to settle. The null and alternative hypotheses associated with Hypothesis I are presented for each model in Table 4.

Section 7b: Hypothesis II

For each of the accused publishers, I also predict that the prices of e-books published by that publisher decreased relative to Random House from any of the time-periods before their settlement date to any of the time-periods after. This means that I predict that the prices of e-books published by Hachette, HarperCollins, and Simon & Schuster decreased from time-period $W$ to any other time-period, the prices of e-books published by Penguin Group decreased from time-period $W$ or $X$ to time-period $Y$ or $Z$, and the prices of e-books published by Macmillan decreased from any prior time-period to time-period $Z$.

This hypothesis checks for the robustness of Hypothesis I by suggesting that prices for firm $A$ were higher in any of its collusive periods compared to any of its competitive time-periods relative to Random House. This hypothesis tests for both the effect of publisher $A$’s settlement and the settlement of the other accused publishers on the price of e-books published by publisher $A$. For this reason, if this hypothesis is proven, it would
suggest three alternative explanations for the change in publisher A’s e-book retail prices due to the settlements of the other publishers. First, the settlement of the other publishers did not affect publisher A’s prices. Second, the settlement of the other publishers also acted to decrease publisher A’s prices. And third, the settlement of the other publishers acted to increase publisher A’s prices but the decreasing effect of publisher A’s settlement on publisher A’s prices was dominant.

The null hypotheses associated with Hypothesis II are presented in Table 5. Isolating the impact of the other publishers’ settlements on publisher A’s prices is looked at next in Hypothesis III.

Section 7c: Hypothesis III

The more difficult case to predict is the movement of an accused publisher's e-book prices between two time-periods before their settlement date or two time-periods after. This should only detect the effect of another publisher's settlement date on this publisher's e-book prices ceteris paribus. For example, publisher B's settlement date may have no effect on publisher A's e-book prices if publisher A has already settled, in which case the difference in publisher A's e-book prices over this time-period relative to Random House should be insignificant from zero (assuming publisher B's settlement date also has no effect on Random House’s prices).

Alternatively, publisher B's settlement date may work to decrease the e-book prices of publisher A if publisher A has not yet settled in two ways. First, publisher B’s settlement with the D.O.J. could serve as a signal to publisher A of the results of its own impending settlement procedure with the D.O.J., so that publisher A will want to nullify their agency contracts prior to settlement to mitigate financial penalties realized in settlement. Second, the success of the accused publishers' agency contracts relied on the most-favored-nation clause enforcing all publishers' adherence to this pricing model. Publisher B’s settlement means publisher B’s e-book retail prices are competitively determined by e-book retailers again. After settlement, publisher B’s e-book prices may therefore undercut the e-book retail prices of publisher A. To compete, publisher A may also reduce their e-book prices
either by setting lower e-book retail prices directly through their agency contracts with their retailers or by abandoning their agency contracts altogether.

In either of the above two cases, the difference in publisher A’s e-book prices over this time-period relative to Random House should be significantly different from zero (specifically in the negative direction) because their e-book retail prices decreased. The null hypotheses associated with Hypothesis III are presented in Table 6.

Section 8: Results

The Stata regression output for Model I can be found in Table 7. The Stata regression output for Model II can be found in Table 8, and the Stata regression output for Model III can be found in Table 9. The movement of e-book retail prices for each of the Big Six publishers over the entire time-period of analysis is graphed in Figure 4. This figure shows the average e-book retail price for each of the six publishers in each of the four time-periods. A separate graph of the movement of the e-book retail prices for each of the accused publishers relative to Random House in each of the four time-periods is provided in Figures 5-9. From these figures we can see that the e-book retail prices for all six publishers were trending downwards over the entire period of study. The question the difference-in-differences analysis sets out to answer is whether the decrease in prices for the accused publishers was of a significantly greater magnitude than the decrease in prices for Random House, the control group.

A summary of all of the coefficient estimates necessary for the difference-in-differences analysis for each model is presented in Table 10. From Table 10, we can see that Model II, where ISBN-level and retailer-level fixed effects are included, seems to be the most explanatory of the three models. Its R-squared value is 0.8422 and its adjusted R-squared value is 0.8392, compared to 0.1917 and 0.1913 respectively in Model I, and 0.5917 and 0.5912 in Model III. The improved explanatory power of Model II needs to be accepted tentatively, however, because Model II includes many more variables than the other two models as ISBN-level fixed effects require 499 additional dummy variables, which artificially increase Model II’s R-squared value. Regardless, Model II is
most similar to Ashenfelter’s (2013) analysis and is therefore used in presenting the results for Figures 4-9 and Tables 15-16.

The actual difference-in-differences estimators can be seen for Models I, II, and III in Tables 11, 12, and 13 respectively. These tables show the estimated difference in prices for all five of the accused publishers between all combinations of the four different time-periods: W, X, Y, and Z relative to the change in prices for Random House over the same time-periods. From this table, I pull out the corresponding D.i.D. coefficient estimates for each of my three hypotheses in Tables 14, 15, and 16 respectively. The results for each of these hypotheses are explained in turn below.

Section 8a: Results for Hypothesis I

Hypothesis I, as detailed in Section 7a and Table 4, predicts that the average e-book retail price for each of the five accused publishers decreased from the time-period immediately prior to their settlement date to the time-period immediately following their settlement date relative to the change in the average e-book retail price for Random House over the same time-periods. The results for Hypothesis I are presented in Table 14 for each model and each of the five accused publishers. These price trends can also be seen in Figure 4.

As can be seen in Table 14, the change in the average e-book retail price for all five publishers except for Macmillan is negative and significant at the one percent level. After each of their respective settlements, Hachette’s prices dropped by an average of 23.1 percent, HarperCollins’ dropped by an average of 23.6 percent, Simon & Schuster’s dropped by an average of 15.5 percent, and Penguin Group’s dropped by an average of 11.6 percent. These averages are calculated by averaging over the three models (columns) in Table 14.

Macmillan is the only publisher that goes against Hypothesis I. Model I and Model III both estimate a positive and significant at the one percent level change in e-book retail prices for e-books published by Macmillan from the period immediately prior to their settlement (Y) to the period immediately following their settlement (Z) relative to Random
House. Model II estimates no significant difference between the change in prices of Macmillan and those of Random House over these time-periods.

I have two possible explanations for why Macmillan serves as a counterexample to Hypothesis I. First, Macmillan was the last of the accused publisher to settle and did so after the verdict had been declared for *U.S. v. Apple Inc. et. al. (2013)*. This means that Macmillan may have dropped their prices prior to their settlement, for example at the settlements of another of the accused publishers. As will be explained further in Section 8b, Tables 11-13 show that Macmillan’s e-book retail prices did decrease (significant at the one percent level) from the time-periods immediately prior to the time-period immediately after the settlement of all the other accused publishers relative to Random House. The only caveat to this is that Model I predicts that the change in Macmillan’s prices was insignificantly different from the change in Random House’s prices after the settlement of Penguin Group. This suggests that Macmillan’s prices may have already come down before their own settlement date, and this is why they do not follow the same trend as the other publishers.

A second explanation for Macmillan’s price movement after their settlement can be seen in Figure 9. Here it is shown that Random House experienced a steep price decline from time-period Y to time-period Z. Figure 9 and Table 10 both show that prices for e-books published by Macmillan did decrease from time-period Y to time-period Z, but this price-decrease was significantly less than the price-decrease for Random House over the same time-periods in Models I and III (significant at the one and five percent levels respectively) and insignificant from the price-decrease for Random House in Model II.

This means that the inconsistency of Macmillan is more a function of Random House’s large price decrease from time-period Y to time-period Z. The question then becomes: why did Random House experience such a large price-decrease at this time? The transition from time-period Y to time-period Z is approximately one month after the verdict of *U.S. v. Apple Inc. et. al. (2013)* and the merger of Penguin Group and Random House. Maybe it really is that the settlement of Macmillan had a dramatic effect on Random House’s prices, or maybe either of these other two factors influenced Random House’s
prices. As will be elaborated on more in Section 9, this is a good opportunity for future study.

We can also look at the average price-change across all five of the accused publishers for each of the three models from immediately before to immediately after each publisher’s settlement date relative to Random House. Table 14 shows that the average price-decrease from the time-period immediately before to the time-period immediately after each publisher’s settlement is 18.5 percent for Model I, 15.8 percent for Model II, and 15.7 percent for Model III. All three of these estimates are significantly different from zero at the one percent level. Additionally, these estimates can be compared to Ashenfelter’s (2013) estimate of a 16.8 percent average increase in e-book retail prices due to the adoption of the accused publishers’ agency contracts. Model I’s estimate of 18.5 percent is actually greater in magnitude than Ashenfelter’s, significant at the one percent level. This means that Model I estimates that the average e-book retail price for the accused publishers has come down further relative to Random House since the accused publishers' settlements than it rose after the adoption of the agency contracts.

Model II’s estimate of a 15.8 percent price-decrease is less than Ashenfelter’s estimate, significant at the five percent level. And Model III’s estimate of a 15.7 percent price-decrease is insignificantly different from Ashenfelter’s estimate. Model II is significantly less than Ashenfelter’s estimate, at the five percent level, while Model III is insignificantly different from his estimate even though the point estimate of Model II is higher (and closer to Ashenfelter’s estimate), because on average Model III has higher standard errors for the difference-in-differences estimators than Model II.

Averaging over these estimates from each of the three models for the average price-change across all five of the accused publishers from immediately before to immediately after each publisher’s settlement date relative to Random House, we get a 16.7 percent price decrease (Table 14). This means that on average, e-book prices for the accused publishers decreased by 16.7 percent from the time-period immediately before to the time-period immediately after each of their respective settlement dates relative to Random House. This estimate is assumedly significantly different from zero, but
insignificantly different from Ashenfelter’s (2013) estimate.

Seen in tandem, these estimates for average price decrease after the accused publishers’ respective dates provide evidence that average e-book retail prices for the five accused publishers relative to Random House have come down to their pre-agency levels. As will be elaborated on more in Section 9, these results provide evidence that the D.O.J. led a successful antitrust intervention in the case of U.S. v. Apple Inc. et. al. (2013).

Section 8b: Results for Hypothesis II

Hypothesis II, as detailed in Section 7b and Table 5, predicts that the average e-book retail price for each of the five accused publishers decreased from any time-period before their settlement date to any time-period after their settlement date relative to the change in the average e-book retail price for Random House over the same time-periods. The results for Hypothesis II are presented in Table 15 for Model II and each of the five accused publishers. These price trends can also be seen in Figure 4.

The change in prices for Hachette, HarperCollins, and Simon & Schuster, the three publisher who were the first to settle with the D.O.J., all confirm Hypothesis II. Their prices were higher in their collusion period ($W$) than in any one of their competition periods ($X$, $Y$, or $Z$) relative to Random House, significant at the one percent level. On average, Hachette’s prices decreased by 20.5 percent, HarperCollins’ prices decreased by 24.9 percent, and Simon & Schuster’s prices decreased by 14.6 percent from time-period $W$ to their competition time-periods.

Penguin Group and Macmillan both only partially confirm Hypothesis II. Table 15 shows that the prices for Penguin Group were higher in their collusion period ($W$ and $X$) than in their first competition period ($Y$) relative to Random House and significant at the one percent level, but were insignificantly different or even significantly lower in their collusion period than in their second competition period ($Z$) relative to Random House. That being said, the average price decrease for Penguin Group from any time-period in their collusion period ($W$ or $X$) to any time-period in their competition period ($Y$ or $Z$) was 5.2 percent.
Similarly, the prices for Macmillan were higher in time-period $W$ and time-period $X$ than in their competition period ($Z$) relative to Random House and significant at the one percent level, but Macmillan’s e-book retail prices were insignificantly different in time-periods $Y$ and $Z$ relative to Random House. The average price decrease for Macmillan from any time-period in their collusion period ($W$, $X$, or $Y$) to their competition period ($Z$) was 7.5 percent.

As was touched upon in Section 8a, the insignificant estimates for the change in e-book prices for Penguin Group and Macmillan relative to Random House could be due to Random House's dramatic price decrease from time-period $Y$ to time-period $Z$. Table 10 shows that prices did decrease for Penguin Group and Macmillan across the above time-periods. Random House's decrease just was significantly greater in magnitude or insignificantly different from this change. This recurring point will be elaborated on further in Section 9.

One of the most interesting questions relating to Hypothesis II is to look at what was the average change in e-book retail prices for all five of the accused publishers relative to Random House over the entire period of my analysis (from period $W$ to period $Z$). This estimate for the change in average e-book retail prices would capture the effects of all five of the accused publishers’ settlements with the D.O.J., the beginning and ending of the trial of *U.S. v. Apple Inc. et. al. (2013)*, as well as the merger of Penguin Group and Random House.

From Table 15, we can see that Model II estimates the change in prices from time-period $W$ to $Z$ as negative and significant at the one percent level for all five of the accused publishers except for Penguin Group. Model II estimates that Penguin Group’s prices had no significant change from time-period $W$ to time-period $Z$. The change in the average e-book retail price for all five of the accused publishers relative to Random House from time-period $W$ to $Z$ is -5.1 percent for Model I, -10.5 percent for Model II, and -7.7 percent for Model III. All three of these estimates are significant at the one percent level. The average change for all three models is -7.8 percent.

Similarly to Hypothesis I, these estimates can be compared to Ashenfelter’s (2013)
estimate of a 16.8 percent increase in average e-book retail prices due to the adoption of the accused publishers’ agency contracts. The three estimates of the change in average e-book retail price for the five accused publishers relative to Random House over the entire period of my analysis (period \( W \) to period \( Z \)) provided by Models I, II, and III are each smaller in magnitude than Ashenfelter’s estimate by an average of 10.8 percentage points, significant at the one percent level. These results suggest that over the entire time-period of my analysis, the average e-book retail prices of the accused publishers did not fall back to their pre-agency levels. This idea will be explored further in Section 9.

Section 8c: Results for Hypothesis III

Hypothesis III, as detailed in Section 7c and Table 6, predicts that the average e-book retail price for each of the five accused publishers decreased from the time-period immediately before to the time-period immediately after the settlement date of each of the other accused publishers relative to the change in the average e-book retail price for Random House over the same time-periods. The results for Hypothesis III are presented in Table 16 for Model II and each of the five accused publishers. These price trends can also be seen in Figure 4.

As can be seen in Table 6, when looking at two time-periods either before or after publisher \( A \)’s settlement date, publisher \( A \)’s publisher dummy variable and collusion-interaction term cancel out in the difference-in-differences analysis. For this reason, all of the accused publishers share the same D.i.D. estimator for a settlement date that is not their own.

Hypothesis III is confirmed for the first and second settlements but not the third (Table 16). Penguin Group and Macmillan’s prices decreased by approximately 4.5 percent (significant at the one percent level) relative to Random House after the joint settlement of Hachette, HarperCollins, and Simon & Schuster. Similarly, the prices of Hachette, HarperCollins, Simon & Schuster, and Macmillan decreased by approximately 7.4 percent (significant at the one percent level) relative to Random House after the settlement of Penguin Group. Conversely, the prices of Hachette, HarperCollins, Simon &
Schuster, and Penguin Group increased by approximately 20.1 percent (significant at the one percent level) relative to Random House after the settlement of Macmillan.

These results are interesting for a couple of reasons. First, it shows that Penguin Group’s and Macmillan’s prices did drop relative to Random House after the first settlement date by about 4.5 percent, significant at the one percent level. This could be seen as evidence that Hachette’s, HarperCollins’, and Simon & Schuster’s settlement acted as signal for Penguin Group and Macmillan that they would also be involved in costly settlements with the D.O.J. Alternatively, after their settlement, Hachette’s, HarperCollins’, and Simon & Schuster’s new competitive prices could have worked to pull down Penguin Group’s and Macmillan’s agency prices.

This trend is also observed in the second settlement. The average e-book retail price for e-books published by Hachette, HarperCollins, Simon & Schuster, and Macmillan decreased relative to Random House after the settlement of Penguin Group by about 7.4 percent, significant at the one percent level. For the most part, this decrease in prices cannot be explained by the signalling hypothesis presented before, because Hachette, HarperCollins, and Simon & Schuster had already settled. Instead, maybe Penguin Group’s new competitive prices continued to drive industry prices down. Either explanation could apply to the decrease in Macmillan’s prices over this time-period.

The fact that the last settlement (Macmillan’s) led to a significant increase in prices for the other accused publishers relative to Random House is difficult to explain. By the time of Macmillan’s settlement, the other accused publishers had already settled with the D.O.J., so it is possible that Macmillan’s settlement could have a small effect on their prices. Additionally, as has been mentioned before, Random House experienced a dramatic decrease in prices from time-period Y to time-period Z (the periods before and after Macmillan’s settlement date). This could be why the difference-in-differences estimators for the change in the prices of the four accused publishers other than Macmillan from immediately before to immediately after Macmillan’s settlement relative to Random House are significantly positive, which contradicts Hypothesis III. The reason why Random House’s prices so dramatically decreased at the time of Macmillan’s settlement is
explored further in Section 9.

Section 9: Conclusion

New York District Judge Denise Cote’s verdict on U.S. v. Apple Inc. et. al. (2013) ruled Apple guilty of price fixing using both a per se approach, citing the anticompetitive nature of the most-favored-nation clause in Apple’s agency contracts, and a rule of reason approach, citing estimates that Apple’s agency contracts resulted in higher prices across the e-book retail market by about 16-19 percent (Ashenfelter, 2013). This paper asks: was this trial and the settlement of the accused publishers an example of a successful antitrust intervention by the U.S. Department of Justice in producing reduced e-book retail prices for consumers? This is an important question to ask, because although all five of the accused publishers settled and Apple was found guilty in U.S. v. Apple Inc. et. al. (2013), the fundamental goal of the Department of Justice’s Antitrust Division and antitrust lawsuits of this nature is to protect consumers from unfair prices. Price fixing eliminates competition on prices among rivals in a market, which works to decrease consumer surplus and societal welfare (Pepall et. al., 2008). This lawsuit was successful, therefore, only if e-book retail prices did in fact decrease after the trial, or more specifically after each publisher settled with the D.O.J. and began re-allowing retailer discounting.

My hypothesis was that e-book prices decreased after the negation of Apple’s agency contracts when final control of e-book retail prices returned to the respective retailers and competition returned to the e-book retail market. I test this hypothesis by measuring the change in e-book prices associated with each of the five accused publishers’ settlements with the D.O.J. This methodology mirrors the work of Ashenfelter (2013), who calculated the increase in prices associated with the introduction of Apple’s agency contracts in April, 2010. This paper estimates the assumed decrease in the accused publishers’ e-book retail prices associated with the dissolution of those contracts across the three aforementioned settlements in 2013. To do this, Ashenfelter and I both use an estimation of cartel overcharge methodology implementing a reduced form model, dummy variable approach (Nieberding, 2006) employing a difference-in-differences
analysis with Random House, the only non-accused publisher among the Big Six, serving as the control group. I depart from Ashenfelter’s analysis by modelling three settlement dates resulting in three collusion dummy variables and four distinct time-periods.

Similar to Ashenfelter’s (2013) study, a significant, negative coefficient on the difference-in-differences estimator for each of the five accused publishers would prove my hypothesis by showing that prices were elevated during the time of collusion, when the accused publishers implemented their agency pricing model on all e-book retailers, and that e-book prices have since come down after the publishers’ settlements with the D.O.J., when the accused publishers’ agency contracts were negated and retailer discounting began.

I find that the average e-book retail prices for all five of the accused publishers decreased from the time-period immediately prior to each publisher’s corresponding settlement date with the D.O.J. to the time-period immediately after relative to Random House over the same time-periods by an average of 16.7 percent (significant at the one percent level). Additionally, I find that average e-book retail prices for all five of the accused publishers decreased by 7.8 percent (significant at the one percent level) relative to Random House over the entire time-period of my analysis.

These results can be compared to Ashenfelter’s (2013) finding of a 16.8 percent increase after the introduction of Apple’s agency contracts. My difference-in-differences estimate of 16.7 percent for the average change in e-book retail prices for each of the accused publishers relative to Random House from the time-period immediately before each publisher’s settlement with the D.O.J. to the time-period immediately following is insignificantly different from Ashenfelter’s (2013) estimate. This suggests that post-agency e-book retail prices have returned to their competitive, pre-agency levels. These results confirm my hypothesis and provide strong evidence for the correctness of the District Court’s verdict, the success of the D.O.J.’s lawsuit as an antitrust intervention, and the misguided intent of Apple to appeal the case.

On the other hand, my difference-in-differences estimate of 7.8 percent for the average change in e-book retail prices for each of the accused publishers relative to
Random House over the entire time-period of my analysis is significantly less than Ashenfelter's (2013) estimate at the one percent level. This suggest that while post-agency e-book retail prices have decreased, they have not fully returned to their competitive, pre-agency levels.

As was explained in Section 8, one of the largest factors driving my insignificant price-change results for the five accused publishers from time-period Y to time-period Z was the dramatic price-decrease in e-books published by Random House over this period, which roughly corresponds to the dates of the trial of U.S. v. Apple Inc. et. al. (2013) and the merger of Penguin Group and Random House. This could bring into question the usefulness of Random House as the control group in my difference-in-differences analysis. For example, it could be that although Random House did not sign agency contracts with any of its retailers and therefore was never accused of price collusion, they could have implemented a profit-maximization strategy to take advantage of the other accused publishers' elevated prices by unilaterally increasing their prices as well.

This would explain why Random House's prices were trending downwards over the entire period of my analysis. As each of the accused publishers settled and their prices were reduced, Random House had to react by also reducing their prices to more competitive levels. This would not invalidate my results, but instead make my case harder to prove, because in the difference-in-differences analysis, the accused publishers prices would have to decrease further in magnitude than the decrease in the prices of e-books published by Random House for their D.i.D. coefficient estimates to be negative. The downward trending prices of Random House and the dramatic decrease in their prices from time-period Y to time-period Z could explain why my estimates for the average decrease in e-book retail prices over the entire time-period of my analysis is significantly smaller in magnitude than the price-increase estimated by Ashenfelter (2013).

Further analysis into the movement of Random House’s prices over this time-period is a good opportunity for future study. Specifically, which factors had the most influence on Random House’s prices from time-period Y to time-period Z: the verdict of U.S. v. Apple Inc. et. al. (2013), their merger with Penguin Group, or the settlement of Macmillan?
It would seem strange for the verdict of *U.S. v. Apple Inc. et. al. (2013)* to have a significant effect on Random House’s prices because Random House was never a defendant or directly involved in the case. It would also seem strange for the merger of Random House with Penguin Group to decrease their prices, because industrial-organization economic theory predicts that mergers generally increase the Herfindahl-Hirschman index of an industry and increase prices (Pepall et. al., 2008).

And lastly, it would seem strange for the settlement of Macmillan to have a significantly greater effect on the e-book prices of Random House compared to the effect of the other accused publishers’ settlements. This may be possible, though, because as Figure 4 shows, Macmillan and Random House are the two publishers that on average have the highest e-book retail prices in the industry. For this reason, Macmillan and Random House could be the closest competitors out of the Big Six, capturing the “higher end” of the market, and it was only after Macmillan settled with the D.O.J. that Random House felt the pressure to decrease their e-book prices, which they had unilaterally increased when Macmillan had signed its agency contracts, to remain competitive with Macmillan. These questions could all be the basis for opportunities of future study.

**Section 10: References**


Connor, J.M. (2003), Private International Cartels: Effectiveness, Welfare, and Anticartel Enforcement, Staff Paper 03-12, West Lafayette, IN: Department of Agricultural Economics, Purdue University.


Morse, B.A. & Hyde, J. (2000), Estimation of Cartel Overcharges: The Case of Archer Daniels Midland and the Market for Lysine, Staff Paper 08-00, West Lafayette, IN: Department of Agricultural Economics, Purdue University.


Section 11: Data Sources

Section 11a: iobyteSolutions.xlsx

The data used in this analysis was received from iobyte Solutions, Inc. (http://www.iobyte.com/), an I.T. consulting company based out of Thornwood, New York. iobyte provides a service called “e-book Market View” for its clients primarily to help publishers analyze “competitive pricing behavior, metadata compliance, author discovery, and merchandising analytics.” My data request was handled by Dan Lubart, a Managing Partner at iobyte Solutions on January 4, 2014.

The data includes 91 weeks of e-book retail pricing data for a sample of 500 best-selling ISBNs for the three biggest e-book retailers: Amazon, Apple, and Barnes & Noble. For each week, I have the price in U.S.D. of these same 500 best-selling e-books for each of these three retailers. Each row of this dataset therefore corresponds to the retail price for an ISBN being sold at one of the aforementioned retailers in a given week.

Section 11b: googleBooks.xls

To supplement the e-book retail pricing data from iobyte Solutions Inc., I queried the Google Books A.P.I. for additional e-book level information. Google Books is a service offered by Google, Inc. allowing users to search and preview millions of books from libraries and publishers all over the world (http://books.google.com/). As part of the service, Google provides the Google Books A.P.I. which allows developers restricted access to Google's database of books serving as the backend to their Google Books service. The script to query the Google Books A.P.I. (https://developers.google.com/books/) was ran on February 18, 19, and 20, 2014.

Using my personal Google/Gmail credentials, I was able to request an A.P.I. key from Google allowing me restricted access to the Google Books A.P.I. (less than 5,000 requests per day). For each unique ISBN from iobyteSolutions.xlsx, I used a JavaScript script to query the Google Books A.P.I. with a given ISBN to retrieve further information such as that ISBN's publication date and genre. This process could be repeated by anyone with Google/Gmail credentials by referring to the Google Books A.P.I. documentation page.
As an e-book's ISBN number is a unique identifier for that e-book, this supplementary Google Books dataset could then be merged with the dataset from iobYTE Solutions Inc. using ISBN as the primary key.
Section 12: Figures

Figure 1
Linear Supply and Demand Graph Demonstrating the Effects of Price Fixing

![Linear Supply and Demand Graph](image)

This figure shows a linear supply and demand graph demonstrating the effects of price fixing. Consider a perfectly-competitive, two-firm industry characterized by the linear supply and demand curves as shown in the graph above. In this industry, the equilibrium quantity sold is $q^*$ and the equilibrium price is $p^*$. Now consider if the two firms formed a cartel and began to collude. This cartel could either reduce their quantity to $q^c$, the cartelized quantity, or just agree to fix their prices at $p^c$, the cartelized price. Either way, we can see from the above graph that the new quantity sold would be $q^c$, where $q^c < q^*$, and the new price would be $p^c$, where $p^c > p^*$.

From this graph, therefore, we can see that as a result of the two firms’ collusion, the industry output decreases ($q^* \rightarrow q^c$), the industry price increase ($p^* \rightarrow p^c$), and there is reduced societal welfare (represented by the newly-formed dead-weight-loss triangle).
Figure 2

Timeline of Relevant Events

2010
January 21    Apple launches its iPad.
April 1       Apple signs agency contracts with the five accused publishers.

2011
--

2012
April 11     The D.O.J. files an antitrust lawsuit against Apple and the five accused publishers.
September 14 Hachette, HarperCollins, and Simon & Schuster settle with the D.O.J.

2013
May 25       Penguin Group settles with the D.O.J.
June 3       The trial of U.S. v. Apple Inc. et. al. (2013) begins.
July 1       Penguin Group and Random House merge.
July 10      Judge Denise Cote delivers her verdict ruling Apple guilty in U.S. v. Apple Inc. et. al. (2013).
August 16    Macmillan settles with the D.O.J.

2014
February 24  Apple files to appeal the verdict of U.S. v. Apple Inc. et. al. (2013).

Above is presented a timeline of all of the significant events relevant to the trial of U.S. v. Apple Inc. et. al. (2013) and this analysis.
Figure 3
The Four Time-Periods Associated with the Three Publisher Settlement Dates

This figure shows the four distinct time-periods that result from the accused publishers' three settlement dates with the D.O.J. Hachette, HarperCollins, and Simon & Schuster settled on September 14, 2012 ($C_{HHS}$). Penguin Group settled on May 25, 2013 ($C_{PG}$), and Macmillan settled on August 16, 2012 ($C_{MM}$).
Figure 4

The Movement of the Average e-book Retail Prices for each of the Big Six Publishers

This figure shows the natural log of the average price of e-books published by each of the Big Six publishers (y-axis) in each of the four time-periods as determined by the five accused publishers’ settlement dates (x-axis). Hachette, HarperCollins, and Simon & Schuster were the first to settle around week 22 of my dataset. Penguin Group was next around week 57. And Macmillan was last to settle around week 68. Random House (in red) serves as the base-line for all of the accused publishers. Note that the above estimates were calculated using Model II.
Figure 5

The Movement of Hachette’s Average e-book Retail Prices Relative to Random House

This figure shows the natural log of the average e-book retail price for Hachette and Random House in each of the four time-periods as defined by the accused publishers’ settlement dates. Hachette settled around week 22. Note that the above estimates were calculated using Model II.
Figure 6

The Movement of HarperCollins’ Average e-book Retail Prices Relative to Random House

This figure shows the natural log of the average e-book retail price for HarperCollins and Random House in each of the four time-periods as defined by the accused publishers’ settlement dates. HarperCollins settled around week 22. Note that the above estimates were calculated using Model II.
This figure shows the natural log of the average e-book retail price for Simon & Schuster and Random House in each of the four time-periods as defined by the accused publishers’ settlement dates. Simon & Schuster settled around week 22. Note that the above estimates were calculated using Model II.
This figure shows the natural log of the average e-book retail price for Penguin Group and Random House in each of the four time-periods as defined by the accused publishers’ settlement dates. Penguin Group settled around week 57. Note that the above estimates were calculated using Model II.
Figure 9

The Movement of Macmillan’s Average e-book Retail Prices Relative to Random House

This figure shows the natural log of the average e-book retail price for Macmillan and Random House in each of the four time-periods as defined by the accused publishers’ settlement dates. Macmillan settled around week 68. Note that the above estimates were calculated using Model II.
Section 13: Tables

Table 1

The Coefficients Affecting Price for each Publisher in each Time-Period

<table>
<thead>
<tr>
<th>W</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{HHS} = 0$</td>
<td>$C_{HHS} = 1$</td>
<td>$C_{HHS} = 1$</td>
<td>$C_{HHS} = 1$</td>
</tr>
<tr>
<td>$C_{PG} = 0$</td>
<td>$C_{PG} = 0$</td>
<td>$C_{PG} = 1$</td>
<td>$C_{PG} = 1$</td>
</tr>
<tr>
<td>$C_{MM} = 0$</td>
<td>$C_{MM} = 0$</td>
<td>$C_{MM} = 0$</td>
<td>$C_{MM} = 0$</td>
</tr>
</tbody>
</table>

HA  $\beta_0 + \beta_2$   $\beta_0 + \beta_1 + \beta_2 + \beta_3$   $\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_{11}$   $\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_{8} + \beta_{11}$

HC  $\beta_0 + \beta_5$   $\beta_0 + \beta_3 + \beta_4 + \beta_5$   $\beta_0 + \beta_3 + \beta_4 + \beta_5 + \beta_{11}$   $\beta_0 + \beta_3 + \beta_4 + \beta_5 + \beta_{8} + \beta_{11}$

MM  $\beta_0 + \beta_7$   $\beta_0 + \beta_3 + \beta_7$   $\beta_0 + \beta_3 + \beta_7 + \beta_{11}$   $\beta_0 + \beta_3 + \beta_6 + \beta_7 + \beta_{8} + \beta_{11}$

PG  $\beta_0 + \beta_{10}$   $\beta_0 + \beta_3 + \beta_{10}$   $\beta_0 + \beta_3 + \beta_9 + \beta_{10} + \beta_{11}$   $\beta_0 + \beta_3 + \beta_8 + \beta_9 + \beta_{10} + \beta_{11}$

SS  $\beta_0 + \beta_{13}$   $\beta_0 + \beta_3 + \beta_{12} + \beta_{13}$   $\beta_0 + \beta_3 + \beta_{11} + \beta_{12} + \beta_{13}$   $\beta_0 + \beta_3 + \beta_8 + \beta_{11} + \beta_{12} + \beta_{13}$

RH  $\beta_0$   $\beta_0 + \beta_3 + \beta_{14}$   $\beta_0 + \beta_3 + \beta_{11} + \beta_{14} + \beta_{16}$   $\beta_0 + \beta_3 + \beta_8 + \beta_{11} + \beta_{14} + \beta_{15} + \beta_{16}$

This table shows the coefficients involved in the difference-in-differences analysis that apply to each publisher in each time-period: $W$, $X$, $Y$, and $Z$. For example, if we are looking at a price observation for e-book $i$ (which is published by Hachette) in week $t$ (which corresponds to time-period $Y$), then all coefficients other than $\beta_0$, $\beta_1$, $\beta_2$, $\beta_3$, and $\beta_{11}$ will have no effect on the price of e-book $i$ because the dummy variables they provide the slope estimate for are zero. This table is the first step in constructing the difference-in-differences analysis used in this study.
Table 2
The First-Difference of the Difference-in-Differences Analysis

<table>
<thead>
<tr>
<th></th>
<th>X - W</th>
<th>Y - W</th>
<th>Z - W</th>
<th>Y - X</th>
<th>Z - X</th>
<th>Z - Y</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HA</strong></td>
<td>(\beta_1 + \beta_3)</td>
<td>(\beta_1 + \beta_3 + \beta_{11})</td>
<td>(\beta_1 + \beta_3 + \beta_8 + \beta_{11})</td>
<td>(\beta_{11})</td>
<td>(\beta_8 + \beta_{11})</td>
<td>(\beta_8)</td>
</tr>
<tr>
<td><strong>HC</strong></td>
<td>(\beta_3 + \beta_4)</td>
<td>(\beta_3 + \beta_4 + \beta_{11})</td>
<td>(\beta_3 + \beta_4 + \beta_8 + \beta_{11})</td>
<td>(\beta_{11})</td>
<td>(\beta_8 + \beta_{11})</td>
<td>(\beta_8)</td>
</tr>
<tr>
<td><strong>MM</strong></td>
<td>(\beta_3)</td>
<td>(\beta_3 + \beta_{11})</td>
<td>(\beta_3 + \beta_6 + \beta_8 + \beta_{11})</td>
<td>(\beta_{11})</td>
<td>(\beta_6 + \beta_8 + \beta_{11})</td>
<td>(\beta_6 + \beta_8)</td>
</tr>
<tr>
<td><strong>PG</strong></td>
<td>(\beta_3)</td>
<td>(\beta_3 + \beta_{9} + \beta_{11})</td>
<td>(\beta_3 + \beta_8 + \beta_9 + \beta_{11})</td>
<td>(\beta_{9} + \beta_{11})</td>
<td>(\beta_8 + \beta_9 + \beta_{11})</td>
<td>(\beta_8)</td>
</tr>
<tr>
<td><strong>SS</strong></td>
<td>(\beta_3 + \beta_{13})</td>
<td>(\beta_3 + \beta_{11} + \beta_{13})</td>
<td>(\beta_3 + \beta_8 + \beta_{11} + \beta_{13})</td>
<td>(\beta_{11})</td>
<td>(\beta_8 + \beta_{11})</td>
<td>(\beta_8)</td>
</tr>
<tr>
<td><strong>RH</strong></td>
<td>(\beta_3 + \beta_{14})</td>
<td>(\beta_3 + \beta_{11} + \beta_{14} + \beta_{16})</td>
<td>(\beta_3 + \beta_8 + \beta_{11} + \beta_{14} + \beta_{15} + \beta_{16})</td>
<td>(\beta_{11} + \beta_{16})</td>
<td>(\beta_{11} + \beta_{15} + \beta_{16})</td>
<td>(\beta_8 + \beta_{15})</td>
</tr>
</tbody>
</table>

This table shows the first-difference change in e-book retail prices for each publisher across each of the possible time-periods. The first difference for each publisher is the difference in e-book prices from one time-period to another. There are four different time-periods, so there are six different differences that can be examined, corresponding to the six columns of this table. For example, row one (HA), column one (X - W) shows the change in average e-book retail price for Hachette from time-period \(W\) to time-period \(X\). Similarly, for all publishers across all time-periods. Calculating the first-differences for each publisher is the second step of the difference-in-differences analysis used in this study.
Table 3

The Difference-in-Differences for each Accused Publisher Relative to Random House

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HC – RH</td>
<td>β4 – β14 – β16</td>
<td>β4 – β14 – β15 – β16</td>
<td>-β16</td>
<td>-β15 – β16</td>
<td>-β15</td>
<td></td>
</tr>
</tbody>
</table>

This table shows the second-difference or difference-in-differences change in e-book retail prices for each accused publisher across each time-period relative to Random House, the base-line. This is the final step of the difference-in-differences analysis used in this study. The first difference for each publisher is the difference in e-book prices from one time-period to another as presented in Table 2. The second difference is that change minus the corresponding change in the e-book prices for Random House across the same time-periods as is presented here.
Table 4

Hypothesis I: Null and Alternative Hypotheses

<table>
<thead>
<tr>
<th></th>
<th>$H_0$</th>
<th>$H_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hachette</td>
<td>$\beta_1 - \beta_{14} = 0$</td>
<td>$\beta_1 - \beta_{14} \neq 0$</td>
</tr>
<tr>
<td>HarperCollins</td>
<td>$\beta_4 - \beta_{14} = 0$</td>
<td>$\beta_4 - \beta_{14} \neq 0$</td>
</tr>
<tr>
<td>Simon &amp; Schuster</td>
<td>$\beta_{12} - \beta_{14} = 0$</td>
<td>$\beta_{12} - \beta_{14} \neq 0$</td>
</tr>
<tr>
<td>Penguin Group</td>
<td>$\beta_9 - \beta_{16} = 0$</td>
<td>$\beta_9 - \beta_{16} \neq 0$</td>
</tr>
<tr>
<td>Macmillan</td>
<td>$\beta_6 - \beta_{15} = 0$</td>
<td>$\beta_6 - \beta_{15} \neq 0$</td>
</tr>
</tbody>
</table>

This table presents the null and alternative hypotheses for each of the three models in Section 6 as laid out in Section 7a, Hypothesis I. The null hypothesis is that the change in e-book prices for each of the accused publishers from the time-period immediately before to immediately after their settlement date is the same as the change in e-book prices for Random House over the same time-periods. The alternative hypothesis is that these two changes are significantly different from each other. In my study, I am particularly interested in the one-sided inequality that the change in e-book prices for the accused publishers after their settlement was greater in magnitude than the change in prices for Random House over the same time-periods, but I test the two-sided inequality because it is more difficult to prove.
This table presents the null hypotheses for each of the three models in Section 6 as laid out in Section 7b, Hypothesis II. The alternative hypotheses are not presented, but are the strict inequality for each entry. For example, the corresponding alternative hypothesis for the first row (HA), first column (X - W) would be $\beta_1 - \beta_{14} \neq 0$. Penguin Group has one additional null and alternative hypothesis compared to the other four accused publishers, because Penguin Group settled in the middle of the other two settlement dates, so unlike the other accused publishers has four distinct combinations of time-period pairs (Y - W, Y - X, Z - W, and Z - X).
Table 6

Hypothesis III: Null Hypotheses

<table>
<thead>
<tr>
<th></th>
<th>$H_0^1$</th>
<th>$H_0^2$</th>
<th>$H_0^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HA, HC, SS Settlement</td>
<td>PG Settlement</td>
<td>MM Settlement</td>
</tr>
<tr>
<td>Hachette</td>
<td>--</td>
<td>$-\beta_{16} = 0$</td>
<td>$-\beta_{15} = 0$</td>
</tr>
<tr>
<td>HarperCollins</td>
<td>--</td>
<td>$-\beta_{16} = 0$</td>
<td>$-\beta_{15} = 0$</td>
</tr>
<tr>
<td>Simon &amp; Schuster</td>
<td>--</td>
<td>$-\beta_{16} = 0$</td>
<td>$-\beta_{15} = 0$</td>
</tr>
<tr>
<td>Penguin Group</td>
<td>$-\beta_{14} = 0$</td>
<td>--</td>
<td>$-\beta_{15} = 0$</td>
</tr>
<tr>
<td>Macmillan</td>
<td>$-\beta_{14} = 0$</td>
<td>$-\beta_{16} = 0$</td>
<td>--</td>
</tr>
</tbody>
</table>

This table presents the null hypotheses for each of the three models in Section 6 as laid out in Section 7c, Hypothesis III. Here, we are trying to detect the cross-effect of accused publisher $B$’s settlement with the D.O.J. on publisher $A$’s e-book prices. Note that because we use a difference-in-differences approach, all other coefficients drop out except for the interaction term between Random House and the collusion dummy variable corresponding to each settlement. As in Table 5, the alternative hypotheses are not presented, but are the strict inequality for each entry. For example, the corresponding alternative hypothesis for the first row (Hachette), second column ($H_1^2$) would be $-\beta_{16} \neq 0$. 
### Table 7

Regression Output for Model I

```stata
. // MODEL I: NO CONTROLS
. reg lprice pha_chhs pha chhs phc_chhs phc pmm_cmm pmm cmm ppg_cpg ppg cpg pss_chhs pss prh_chhs prh_cmm prh_cpg
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 27333</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1197.91277</td>
<td>16</td>
<td>74.8695479</td>
<td>F( 16, 27316) = 404.97</td>
</tr>
<tr>
<td>Residual</td>
<td>5050.06975 27316</td>
<td>.184875888</td>
<td>R² = 0.1917</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>6247.98252 27332</td>
<td>.228595877</td>
<td>Root MSE = 0.42997</td>
<td></td>
</tr>
</tbody>
</table>

| lprice      | Coef.   | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|-------------|---------|-----------|-------|-------|---------------------|
| pha_chhs    | -.1706814 | .0190409 | -8.96 | 0.000 | -.2080024 -.133604 |
| pha         | -.0525725 | .0157591 | -3.34 | 0.001 | -.0834612 -.0216838 |
| chhs        | -.0367231 | .0128435 | -2.86 | 0.004 | -.0618971 -.0115491 |
| phc_chhs    | -.2389773 | .023201 | -10.30 | 0.000 | -.2844525 -.1935021 |
| phc         | -.3073718 | .0196399 | -15.65 | 0.000 | -.3548671 -.2688765 |
| pmm_cmm     | -.1011743 | .0256238 | -3.95 | 0.000 | -.1513983 -.0509503 |
| pmm         | -.1345187 | .0169508 | -7.94 | 0.000 | -.1677432 -.1012942 |
| cmm         | -.0710581 | .010468 | -6.79 | 0.000 | -.0915759 -.0505404 |
| ppg_cpg     | -.1477632 | .0147868 | -9.99 | 0.000 | -.1767461 -.1187803 |
| ppg         | .0996505  | .0149981 | 6.64  | 0.000 | .0702536 .1290475 |
| cpg         | .0372308  | .0106299 | -3.50 | 0.000 | -.0580659 -.0163957 |
| pss_chhs    | -.3379926 | .023504 | -14.38 | 0.000 | -.3840616 -.2919236 |
| pss         | .1168604  | .0189581 | 6.16  | 0.000 | .0797016 .1540192 |
| prh_chhs    | .0658378  | .0182407 | 3.61  | 0.000 | .0300852 .1015905 |
| prh_cmm     | -.2200219 | .0194725 | -11.30 | 0.000 | -.2581891 -.1818548 |
| prh_cpg     | .0082479  | .0188256 | 0.44  | 0.661 | -.0286514 .0451471 |
| _cons       | 2.412866  | .0099324 | 242.93 | 0.000 | .2393398 .2432334 |

This table shows the Stata regression output for Model I. This model only includes the essential variables for the difference-in-differences analysis: the publisher dummy variables, the collusion dummy variables, and the interaction terms between each of the publishers and their corresponding collusion variable. Note that by including publisher dummy variables, this model implements publisher fixed effects with Random House serving as the omitted state.
Regression Output for Model II

This table shows the Stata regression output for Model II. This model includes all of the variables included in Model I plus ISBN-level and retailer-level fixed effects. The coefficient estimates for the 499 included ISBN dummy variables are not shown. The e-book with the lowest ISBN number serves as the omitted state.
This table shows the Stata regression output for Model III. This model includes all of the variables from Model II except implements a list of ISBN-level controls instead of ISBN-level fixed effects. Refer to the Data Appendix or Section 6c for more information on each of the included controls.
Table 10

Regression Results for all Models

<table>
<thead>
<tr>
<th></th>
<th>MODEL I</th>
<th>MODEL II</th>
<th>MODEL III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NO CONTROLS</td>
<td>ISBN, RETAILER FE</td>
<td>ISBN CONTROLS</td>
</tr>
<tr>
<td>HA*CHHS</td>
<td>-.1706814*** (.0190409)</td>
<td>-.1949486*** (.0095685)</td>
<td>-.1981623*** (.0144512)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HC*CHHS</td>
<td>-.2389773*** (.023201)</td>
<td>-.2669689*** (.0118233)</td>
<td>-.1212092*** (.0171493)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MM*CMM</td>
<td>-.1011743*** (.0256238)</td>
<td>-.2048251*** (.0140848)</td>
<td>-.1685139*** (.01896)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PG*CPG</td>
<td>-.1477632*** (.0147868)</td>
<td>-.0455045*** (.0078153)</td>
<td>-.0363178*** (.0111682)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS*CSS</td>
<td>-.3379926*** (.023504)</td>
<td>-.1280232*** (.0125442)</td>
<td>-.1775614*** (.0173349)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RH*CHHS</td>
<td>.0658378*** (.0182407)</td>
<td>.0437479*** (.0090987)</td>
<td>.0972207*** (.0138277)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RH*CMM</td>
<td>-.2200219*** (.0194725)</td>
<td>-.1792447*** (.0091111)</td>
<td>-.2139291*** (.0147595)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RH*CPG</td>
<td>.0082479 (.0188256)</td>
<td>.0794605*** (.0091004)</td>
<td>.0592005*** (.0143266)</td>
</tr>
</tbody>
</table>

N 27,333  27,333  23,740
F-statistic 0.0000  0.0000  0.0000
R-squared 0.1917  0.8422  0.5917
Adj. r-squared 0.1913  0.8392  0.5912

*** p-value <= 0.01  ** p-value <= 0.05  * p-value <= 0.1

This table shows the regression results for all three models. The stars represent the significance level of the corresponding p-values for each coefficient estimate and the standard errors are presented in parentheses. Included are only the coefficients on the interaction terms between the publishers and their respective collusion dummy variable, because these are the only values needed in filling out Table 3 with the corresponding difference-in-differences estimates for each of the accused publishers as is done in Tables 11, 12, and 13.
Table 11

D.i.D. Linear Combinations of Estimators for Model I

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HA</td>
<td>-.2436921*** (-21.627%)</td>
<td>-.2386383*** (-21.230%)</td>
<td>-.0264859 (-2.614%)</td>
<td>.0050538 (.507%)</td>
<td>.2172062*** (24.260%)</td>
<td>.2121524*** (23.634%)</td>
</tr>
<tr>
<td>HC</td>
<td>-.2908408*** (-25.237%)</td>
<td>-.285787*** (-24.858%)</td>
<td>-.0736345*** (-7.099%)</td>
<td>.0050538 (.507%)</td>
<td>.2172062*** (24.260%)</td>
<td>.2121524*** (23.634%)</td>
</tr>
<tr>
<td>MM</td>
<td>-.0647916*** (-6.274%)</td>
<td>-.0597378*** (-5.799%)</td>
<td>.061614** (6.355%)</td>
<td>.0050538 (.507%)</td>
<td>.1264056*** (13.474%)</td>
<td>.1213518*** (12.902%)</td>
</tr>
<tr>
<td>PG</td>
<td>-.0647916*** (-6.274%)</td>
<td>-.2202814*** (-19.771%)</td>
<td>-.008129 (-.810%)</td>
<td>-.1554898*** (-14.400%)</td>
<td>.0566626*** (5.830%)</td>
<td>.2121524*** (23.634%)</td>
</tr>
<tr>
<td>SS</td>
<td>-.4052216*** (-33.317%)</td>
<td>-.4001678*** (-32.979%)</td>
<td>-.1880153*** (-17.140%)</td>
<td>.0050538 (.507%)</td>
<td>.2172062*** (24.260%)</td>
<td>.2121524*** (23.634%)</td>
</tr>
</tbody>
</table>

*** p-value <= 0.01 ** p-value <= 0.05 * p-value <= 0.1

This table shows the difference-in-differences, linear combination results for Model I. The stars represent the significance level of the corresponding p-values for each entry and the percentage change for each difference-in-differences estimator is presented in parentheses. This table fills out the linear combinations of coefficients laid out in Table 3 with the regression results from Table 10, column one. The null hypothesis here is that the linear combination of coefficients for each entry is equal to zero, against the alternative hypothesis that the linear combination is not equal to zero. See Section 7 for a more in-depth explanation of my hypotheses.
### Table 12

D.i.D. Linear Combination of Estimators for Model II

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HA</td>
<td>-.242142***</td>
<td>-.3192902***</td>
<td>-.1361953***</td>
<td>-.0771482***</td>
<td>.1059466***</td>
<td>.1830949***</td>
</tr>
<tr>
<td></td>
<td>(-21.506%)</td>
<td>(-27.334%)</td>
<td>(-12.733%)</td>
<td>(-7.425%)</td>
<td>(11.176%)</td>
<td>(20.093%)</td>
</tr>
<tr>
<td>HC</td>
<td>-.2981178***</td>
<td>-.375266***</td>
<td>-.1921711***</td>
<td>-.0771482***</td>
<td>.1059466***</td>
<td>.1830949***</td>
</tr>
<tr>
<td></td>
<td>(-25.779%)</td>
<td>(-31.289%)</td>
<td>(-17.483%)</td>
<td>(-7.425%)</td>
<td>(11.176%)</td>
<td>(20.093%)</td>
</tr>
<tr>
<td>MM</td>
<td>-.0456315***</td>
<td>-.1227797***</td>
<td>-.1358809***</td>
<td>-.0771482***</td>
<td>-.0902495***</td>
<td>-.0131012</td>
</tr>
<tr>
<td></td>
<td>(-4.461%)</td>
<td>(-11.554%)</td>
<td>(-12.705%)</td>
<td>(-7.425%)</td>
<td>(-8.630%)</td>
<td>(-1.302%)</td>
</tr>
<tr>
<td>PG</td>
<td>-.0456315***</td>
<td>-.1724427***</td>
<td>.0106522</td>
<td>-.1268112***</td>
<td>.0562837***</td>
<td>.1830949***</td>
</tr>
<tr>
<td></td>
<td>(-4.461%)</td>
<td>(-15.839%)</td>
<td>(1.071%)</td>
<td>(-11.910%)</td>
<td>(5.790%)</td>
<td>(20.093%)</td>
</tr>
<tr>
<td>SS</td>
<td>-.1698219***</td>
<td>-.2469701***</td>
<td>-.0638753***</td>
<td>-.0771482***</td>
<td>.1059466***</td>
<td>.1830949***</td>
</tr>
<tr>
<td></td>
<td>(-15.618%)</td>
<td>(-21.884%)</td>
<td>(-6.188%)</td>
<td>(-7.425%)</td>
<td>(11.176%)</td>
<td>(20.093%)</td>
</tr>
</tbody>
</table>

*** p-value <= 0.01  ** p-value <= 0.05  * p-value <= 0.1

This table shows the difference-in-differences, linear combination results for Model II. The stars represent the significance level of the corresponding p-values for each entry and the percentage change for each difference-in-differences estimator is presented in parentheses. This table fills out the linear combinations of coefficients laid out in Table 3 with the regression results from Table 10, column two. The null hypothesis here is that the linear combination of coefficients for each entry is equal to zero, against the alternative hypothesis that the linear combination is not equal to zero. See Section 7 for a more in-depth explanation of my hypotheses.
Table 13

D.i.D. Linear Combinations of Estimators for Model III

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HA</strong></td>
<td>-.3033566***</td>
<td>-.3422615***</td>
<td>-.1315229***</td>
<td>-.0389049***</td>
<td>.1718337***</td>
<td>.2107386***</td>
</tr>
<tr>
<td></td>
<td>(-26.166%)</td>
<td>(-28.984%)</td>
<td>(-12.324%)</td>
<td>(-3.816%)</td>
<td>(18.748%)</td>
<td>(23.459%)</td>
</tr>
<tr>
<td><strong>HC</strong></td>
<td>-.2192226***</td>
<td>-.2581275***</td>
<td>-.0473889***</td>
<td>-.0389049***</td>
<td>.1718337***</td>
<td>.2107386***</td>
</tr>
<tr>
<td></td>
<td>(-19.686%)</td>
<td>(-22.750%)</td>
<td>(-4.628%)</td>
<td>(-3.816%)</td>
<td>(18.748%)</td>
<td>(23.459%)</td>
</tr>
<tr>
<td><strong>MM</strong></td>
<td>-.0985713***</td>
<td>-.1374762***</td>
<td>-.0897204***</td>
<td>-.0389049***</td>
<td>.0088509</td>
<td>.0477558**</td>
</tr>
<tr>
<td></td>
<td>(-9.387%)</td>
<td>(-12.844%)</td>
<td>(-8.581%)</td>
<td>(-3.816%)</td>
<td>(.889%)</td>
<td>(4.891%)</td>
</tr>
<tr>
<td><strong>PG</strong></td>
<td>-.0985713***</td>
<td>-.1877957***</td>
<td>.0229429</td>
<td>-.0892244***</td>
<td>.1215142***</td>
<td>.2107386***</td>
</tr>
<tr>
<td></td>
<td>(-9.387%)</td>
<td>(-17.122%)</td>
<td>(2.321%)</td>
<td>(-8.536%)</td>
<td>(12.921%)</td>
<td>(23.459%)</td>
</tr>
<tr>
<td><strong>SS</strong></td>
<td>-.285704***</td>
<td>-.3246089***</td>
<td>-.1138703***</td>
<td>-.0389049***</td>
<td>.1718337***</td>
<td>.2107386***</td>
</tr>
<tr>
<td></td>
<td>(-24.851%)</td>
<td>(-27.719%)</td>
<td>(-10.763%)</td>
<td>(-3.816%)</td>
<td>(18.748%)</td>
<td>(23.459%)</td>
</tr>
</tbody>
</table>

*** p-value <= 0.01 ** p-value <= 0.05 * p-value <= 0.1

This table shows the difference-in-differences, linear combination results for Model III. The stars represent the significance level of the corresponding p-values for each entry and the percentage change for each difference-in-differences estimator is presented in parentheses. This table fills out the linear combinations of coefficients laid out in Table 3 with the regression results from Table 10, column three. The null hypothesis here is that this linear combination of coefficients for each entry is equal to zero, against the alternative hypothesis that the linear combination is not equal to zero. See Section 7 for a more in-depth explanation of my hypotheses.
Table 14

Results for Hypothesis I

<table>
<thead>
<tr>
<th></th>
<th>Model I (%)</th>
<th>Model II (%)</th>
<th>Model III (%)</th>
<th>Average (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Macmillan</strong></td>
<td>12.902***</td>
<td>-1.302</td>
<td>4.891**</td>
<td>5.497</td>
</tr>
</tbody>
</table>

*** p-value <= 0.01  ** p-value <= 0.05  * p-value <= 0.1

This table shows the results relating to Hypothesis I as laid out in Table 4. For each of the five accused publishers, this table presents the percent change in average e-book retail prices from the period immediately before their settlement date to the period immediately following. The stars for Models I, II, and III represent the significance level of the corresponding p-values for each entry. The last row shows the average percent change in prices for the five accused publishers. Note that significance stars are omitted from the last column because only the point estimates were saved across different regressions and could be used in this calculation of averages.
**Table 15**

Model II Results for Hypothesis II

<table>
<thead>
<tr>
<th></th>
<th>X – W (%)</th>
<th>Y – W (%)</th>
<th>Z – W (%)</th>
<th>Y – X (%)</th>
<th>Z – X (%)</th>
<th>Z – Y (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hachette</td>
<td>-21.506***</td>
<td>-27.334***</td>
<td>-12.733***</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>HarperCollins</td>
<td>-25.779***</td>
<td>-31.289***</td>
<td>-17.483***</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Macmillan</td>
<td>--</td>
<td>--</td>
<td>-12.705***</td>
<td>--</td>
<td>-8.630***</td>
<td>-1.302***</td>
</tr>
<tr>
<td>Penguin Group</td>
<td>--</td>
<td>-15.839***</td>
<td>1.071</td>
<td>-11.910***</td>
<td>5.790***</td>
<td>--</td>
</tr>
<tr>
<td>Simon &amp; Schuster</td>
<td>-15.618***</td>
<td>-21.884***</td>
<td>-6.188***</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

*** p-value <= 0.01  ** p-value <= 0.05  * p-value <= 0.1

This table shows the results relating to Hypothesis II as laid out in Table 5. For each of the five accused publishers, this table presents the percent change in average e-book retail prices from all periods prior to their settlement date to all periods following. Entries are omitted if both time-periods occur either before or after publisher A’s settlement date. For example, in row one both time-periods in the last three columns occur after Hachette’s settlement date, so the values are omitted. The stars represent the significance level of the corresponding p-values for each entry. Note that the estimates provided in this table are calculated using Model II.
### Table 16

Model II Results for Hypothesis III

<table>
<thead>
<tr>
<th></th>
<th>HA, HC, SS Settlement (%)</th>
<th>PG Settlement (%)</th>
<th>MM Settlement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hachette</td>
<td>--</td>
<td>-7.425***</td>
<td>20.093***</td>
</tr>
<tr>
<td>HarperCollins</td>
<td>--</td>
<td>-7.425***</td>
<td>20.093***</td>
</tr>
<tr>
<td>Simon &amp; Schuster</td>
<td>--</td>
<td>-7.425***</td>
<td>20.093***</td>
</tr>
<tr>
<td>Penguin Group</td>
<td>-4.461***</td>
<td>--</td>
<td>20.093***</td>
</tr>
<tr>
<td>Macmillan</td>
<td>-4.461***</td>
<td>-7.425***</td>
<td>--</td>
</tr>
</tbody>
</table>

*** p-value <= 0.01  ** p-value <= 0.05  * p-value <= 0.1

This table shows the results relating to Hypothesis III as laid out in Table 6. For each of the five accused publishers, this table presents the percent change in average e-book retail prices for publisher A from the period immediately before the settlement date of publisher B to the period immediately following. Only the cross-effect of publisher B’s settlement on publisher A’s prices is recorded, so there are omitted values in the column corresponding to publisher A’s settlement date. The stars represent the significance level of the corresponding p-values for each entry. Note that the estimates provided in this table are calculated using Model II.
Section 14: Appendix A -- Calculation of Elasticities

To calculate the percentage change from the difference-in-differences coefficient estimates in Tables 11-13, I use the following procedure. For illustration purposes, I use Hachette as an example, but this methodology directly mirrors that used for any of the other accused publisher, time-period pairs.

Consider that each entry of Table 3 represents the difference-in-differences coefficients estimating the change in average e-book retail price for accused publisher A from one time-period to another relative to Random House. For example, the first row, first column of Table 3 shows that the change in average e-book retail price for e-books published by Hachette from time-period W to time-period X is estimated by $\beta_1 - \beta_{14}$. In functional form, this is presented in Equation 1 below. The change in average retail price for Hachette can really be represented as the difference in the average retail price for Hachette (minus Random House) in time-period X minus its average retail price in time-period W as shown in Equation 2. Using a property of logarithms, this difference of logarithms can be combined in Equation 3. I exponentiate both sides in Equation 4. Equations 5, 6, and 7 show the steps required to transform this difference to give the percent change in Hachette’s prices in terms of W as Hachette transitioned into time-period X.

It can similarly be shown that each entry of Table 3 can be substituted in for $\beta_1 - \beta_{14}$ in Equation 7 to give the percentage change in e-book prices for any of the accused publishers, for any two time-period pairs.

\[
\begin{align*}
\Delta \ln P_{W \rightarrow X}^{HA} &= \beta_1 - \beta_{14} \\
\ln P_X^{HA} - \ln P_W^{HA} &= \beta_1 - \beta_{14} \\
\ln(P_X^{HA} / P_W^{HA}) &= \beta_1 - \beta_{14} \\
P_X^{HA} / P_W^{HA} &= e^{(\beta_1 - \beta_{14})} \\
P_X^{HA} / P_W^{HA} - 1 &= e^{(\beta_1 - \beta_{14})} - 1 \\
(P_X^{HA} - P_X^{HA}) / P_W^{HA} &= e^{(\beta_1 - \beta_{14})} - 1 \\
100 \times (P_X^{HA} - P_X^{HA}) / P_W^{HA} &= 100 \times (e^{(\beta_1 - \beta_{14})} - 1)
\end{align*}
\]
Section 15: Appendix B -- Data Appendix


Each row of my dataset (corresponding to one unit of observation) is composed of an ISBN \(i\), retailer \(j\), week \(t\) triple. For example, one row could show that ISBN 9780061738098 was selling for $9.99 at Barnes & Noble during week 14 (corresponding to an actual date). For clarity purposes, therefore, I describe the variables below in terms of the price of e-book \(i\), with retailer \(j\), in week \(t\).

Section 15a: Overview of Variables

Variable Name: ISBN  
Original Name: ISBN13, isbn13  
Data Source: iobyteSolutions.csv, googleBooks.csv  
Variable description: The 13-digit ISBN number for e-book \(i\).

Descriptive statistics:

Unique values: 500
Variable Name: RETAILER
Original Name: ListID
Data Source: iobyteSolutions.csv
Variable description: The retailer $j$ at which e-book $i$ is being sold for the given price during week $t$.

Descriptive statistics:

<table>
<thead>
<tr>
<th>Retailer</th>
<th>Freq.</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon</td>
<td>5,301</td>
<td>19.39</td>
</tr>
<tr>
<td>Apple</td>
<td>1,864</td>
<td>6.82</td>
</tr>
<tr>
<td>Barnes &amp; Noble</td>
<td>20,168</td>
<td>73.79</td>
</tr>
<tr>
<td>Total</td>
<td>27,333</td>
<td>100.00</td>
</tr>
</tbody>
</table>
Variable Name: PRICE
Original Name: price
Data Source: iobyteSolutions.csv

Descriptive statistics:

<table>
<thead>
<tr>
<th>variable</th>
<th>mean</th>
<th>sd</th>
<th>min</th>
<th>p25</th>
<th>p50</th>
<th>p75</th>
<th>max</th>
</tr>
</thead>
</table>

---

![Histogram of Price Distribution]
Variable Name: \textit{LPRICE}
Original Name: \textit{price}
Data Source: iobyteSolutions.csv
Variable description: The natural log of one plus \textit{PRICE}.

Descriptive statistics:

<table>
<thead>
<tr>
<th>variable</th>
<th>mean</th>
<th>sd</th>
<th>min</th>
<th>p25</th>
<th>p50</th>
<th>p75</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>lprice</td>
<td>2.2348</td>
<td>.4781</td>
<td>0</td>
<td>2.0782</td>
<td>2.3969</td>
<td>2.5642</td>
<td>3.8284</td>
</tr>
</tbody>
</table>

![Histogram of LNP of Price](image)
Variable Name: *PUBLISHER*
Original Name: *fullname*
Data Source: *iobyteSolutions.csv*
Variable description: The publisher of e-book $i$.

Descriptive statistics:

<table>
<thead>
<tr>
<th>Publisher</th>
<th>Freq.</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hachette</td>
<td>6,252</td>
<td>22.87</td>
</tr>
<tr>
<td>HarperCollins</td>
<td>3,178</td>
<td>11.63</td>
</tr>
<tr>
<td>MacMillan</td>
<td>1,943</td>
<td>7.11</td>
</tr>
<tr>
<td>Penguin</td>
<td>5,905</td>
<td>21.60</td>
</tr>
<tr>
<td>Random House</td>
<td>7,522</td>
<td>27.52</td>
</tr>
<tr>
<td>Simon &amp; Schuster</td>
<td>2,533</td>
<td>9.27</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>27,333</td>
<td>100.00</td>
</tr>
</tbody>
</table>
Variable Name: *RANK*
Original Name: *ranking*
Data Source: iobyteSolutions.csv

**Variable description:** The ordinal sales rank for e-book $i$, with retailer $j$, in week $t$.

**Descriptive statistics:**

<table>
<thead>
<tr>
<th>variable</th>
<th>mean</th>
<th>sd</th>
<th>min</th>
<th>p25</th>
<th>p50</th>
<th>p75</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>rank</td>
<td>401.34</td>
<td>911.50</td>
<td>1</td>
<td>80</td>
<td>268</td>
<td>549</td>
<td>49622</td>
</tr>
</tbody>
</table>

---

![Graph showing the distribution of sales ranks.](image)
Variable Name: $LRANK$
Original Name: ranking
Data Source: iobyteSolutions.csv
Variable description: The natural log of RANK.

Descriptive statistics:

<table>
<thead>
<tr>
<th>variable</th>
<th>mean</th>
<th>sd</th>
<th>min</th>
<th>p25</th>
<th>p50</th>
<th>p75</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>lrank</td>
<td>5.23797</td>
<td>1.41287</td>
<td>0</td>
<td>4.382027</td>
<td>5.590987</td>
<td>6.302619</td>
<td>10.81219</td>
</tr>
</tbody>
</table>
Variable Name: AGE
Original Name: date, publishedDate
Data Source: iobyteSolutions.csv, googleBooks.csv
Variable description: The age of e-book $i$, calculated as the date of observation for the price of e-book $i$ at retailer $j$ minus the publication date for e-book $i$.

Descriptive statistics:

<table>
<thead>
<tr>
<th>variable</th>
<th>mean</th>
<th>sd</th>
<th>min</th>
<th>p25</th>
<th>p50</th>
<th>p75</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>792.9826</td>
<td>1049.927</td>
<td>2</td>
<td>103</td>
<td>362</td>
<td>1027</td>
<td>5422</td>
</tr>
</tbody>
</table>

![Histogram of Age of e-book](image-url)
Variable Name: LAGE
Original Name: date, publishedDate
Data Source: iobyteSolutions.csv, googleBooks.csv
Variable description: The natural log of one plus AGE.

Descriptive statistics:

<table>
<thead>
<tr>
<th>variable</th>
<th>mean</th>
<th>sd</th>
<th>min</th>
<th>p25</th>
<th>p50</th>
<th>p75</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>lage</td>
<td>5.721951</td>
<td>1.582505</td>
<td>.6931472</td>
<td>4.634729</td>
<td>5.891644</td>
<td>6.94119</td>
<td>8.59822</td>
</tr>
</tbody>
</table>

![Histogram of LNAge of e-book](image)
Variable Name: GENRE
Original Name: categories
Data Source: googleBooks.csv
Variable description: The literary genre for e-book $i$.

Descriptive statistics:

<table>
<thead>
<tr>
<th>Genre</th>
<th>Freq.</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fiction</td>
<td>21,439</td>
<td>78.44</td>
</tr>
<tr>
<td>Juvenile Fiction</td>
<td>3,590</td>
<td>13.13</td>
</tr>
<tr>
<td>Other</td>
<td>2,304</td>
<td>8.43</td>
</tr>
<tr>
<td>Total</td>
<td>27,333</td>
<td>100.00</td>
</tr>
</tbody>
</table>
Variable Name: RATING
Original Name: averageRating
Data Source: googleBooks.csv
Variable description: The average rating for e-book $i$ based on online ratings by registered Google Books users.

Descriptive statistics:

<table>
<thead>
<tr>
<th>variable</th>
<th>mean</th>
<th>sd</th>
<th>min</th>
<th>p25</th>
<th>p50</th>
<th>p75</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>rating</td>
<td>3.690274</td>
<td>0.4537194</td>
<td>2</td>
<td>3.5</td>
<td>3.5</td>
<td></td>
<td>4.5</td>
</tr>
</tbody>
</table>

![Histogram of Average Rating](image)
Variable Name: RATINGS_COUNT
Original Name: ratingsCount
Data Source: googleBooks.csv
Variable description: The number of ratings for e-book i used in calculating RATING.

Descriptive statistics:

<table>
<thead>
<tr>
<th>variable</th>
<th>mean</th>
<th>sd</th>
<th>min</th>
<th>p25</th>
<th>p50</th>
<th>p75</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>ratingscount</td>
<td>747.4792</td>
<td>1268.803</td>
<td>1</td>
<td>47</td>
<td>145</td>
<td>809</td>
<td>7499</td>
</tr>
</tbody>
</table>

![Histogram of Ratings Count](image-url)
Variable Name: LRATINGSCOUNT  
Original Name: ratingsCount  
Data Source: googleBooks.csv  
Variable description: The natural log of one plus RATINGSCOUNT.

Descriptive statistics:

<table>
<thead>
<tr>
<th>variable</th>
<th>mean</th>
<th>sd</th>
<th>min</th>
<th>p25</th>
<th>p50</th>
<th>p75</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>lratingsco-t</td>
<td>5.214898</td>
<td>1.801852</td>
<td>.6931472</td>
<td>3.871201</td>
<td>4.983607</td>
<td>6.697034</td>
<td>8.922658</td>
</tr>
</tbody>
</table>

---

![Histogram of LRATINGSCOUNT](image)
Variable Name: PAGECOUNT
Original Name: pageCount
Data Source: googleBooks.csv
Variable description: The page count for e-book i in its corresponding print version.

Descriptive statistics:

<table>
<thead>
<tr>
<th>variable</th>
<th>mean</th>
<th>sd</th>
<th>min</th>
<th>p25</th>
<th>p50</th>
<th>p75</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>pagecount</td>
<td>435.526</td>
<td>353.9421</td>
<td>15</td>
<td>320</td>
<td>384</td>
<td>480</td>
<td>5216</td>
</tr>
</tbody>
</table>

![Histogram of page count](image.png)
**Variable Name:** LPAGECOUNT  
**Original Name:** pageCount  
**Data Source:** googleBooks.csv  
**Variable description:** The natural log of PAGECOUNT.

### Descriptive statistics:

<table>
<thead>
<tr>
<th>variable</th>
<th>mean</th>
<th>sd</th>
<th>min</th>
<th>p25</th>
<th>p50</th>
<th>p75</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>lpagecount</td>
<td>5.927828</td>
<td>.5710497</td>
<td>2.70805</td>
<td>5.768321</td>
<td>5.950643</td>
<td>6.173786</td>
<td>8.559486</td>
</tr>
</tbody>
</table>

![Histogram of lpagecount](image-url)