File-Sharing and Film Revenues: Is there a Displacement Effect?*

Jordi McKenzie[†]
Macquarie University

W. D. Walls
University of Calgary

April 24, 2014

Abstract

This study examines the impact of peer-to-peer (P2P) file-sharing on the Australian theatrical film industry. Using a large data set of torrent downloads observed on three popular P2P networks, we find evidence of a sales displacement effect on box office revenues. However, although statistically significant, the economic significance of this displacement appears relatively small. To establish causality, we propose a downloading cost function which relates download cost to a proxy for the number of peers in the download swarm and download file size. As a further determinant of downloading cost, we also make use of substantial structural progression within internet service provider industry over our sample period which resulted in large increase in internet subscriptions, speeds and downloads capacities. We observe that the release gap between the US and Australian markets is a key contributor to piracy early in a film's theatrical life; this finding provides a partial explanation for the industry's move toward coordinated worldwide releases.

*We are grateful to participants of the University of Sydney Microeconometric and Public Policy working group (Sydney, October 2012), Motion Picture Scholars Workshop (Los Angeles, November 2012), University of Melbourne Institute of Applied Economic and Social Research seminar series (Melbourne, February 2013), 42nd Australian Conference of Economists (Perth, July 2013), Econometric Society Australasian Meeting 2013 (Sydney, July 2013), and Macquarie University seminar series (Sydney, August 2013) for comments that have helped to improve the manuscript. Any remaining errors are our own.

[†] *Corresponding author:* Faculty of Business and Economics, Macquarie University, Sydney, Australia. Email: `jordi.mckenzie@mq.edu.au`.

1 Introduction

Since the arrival of peer-to-peer (P2P) file-sharing technologies more than a decade ago, intellectual property protection has become an increasingly important issue for copyright protected products that can be digitally reproduced. P2P file-sharing technologies, such as BitTorrent, allow internet users to illegally share copyrighted content (e.g. software, games, books, music, television shows and films) at minimal cost provided they have a high-speed internet connection and sufficient download capacity. The increasing popularity of such services, particularly among younger generations, has created something of a revolution to the way in which many people access and consume content. In response to the increasing incidence of 'digital piracy', owners of copyrighted content have pursued legal recourse in many countries, with high-profile cases brought against file-sharing services, individuals using file-sharing services, and internet service providers (ISPs).

Although it is extremely difficult to put an exact figure on the extent and costs of digital piracy, a recent annual industry survey by Business Software Alliance of 15,000 computer users across 33 countries found that more than 57% of respondents admitted to pirating software in 2012, up from 42% in 2011. Another recent study of piracy habits in the US found that 46% of adults have bought, copied, or downloaded unauthorised music, TV shows or films and that these practices correlate strongly with youth and moderately with higher incomes.[1] In Europe, a 2010 study by the International Chamber of Commerce found that internet pirates downloaded € 10 billion worth of music, film and television and claimed that digital piracy could cost the content industries € 240 billion in revenue and 1.2 million jobs by 2015. And in Australia, two recent (2011) studies by Australian Content Industry Group (ACIG) and Australian Federation Against Copyright Theft (AFACT) estimate annual losses at A$900m and A$1.37b, respectively.[2]

Although numerous industry studies have estimated significant costs from digital piracy, many of them likely overestimate these costs because calculations typically assume a one-for-one sales displacement effect from illegal downloads. This implies that everyone who downloaded the product would have been prepared to pay full price in the absence of the illegal alternative—an extremely uncomfortable assumption. More measured industry studies have assumed some 'substitution rate' between number of downloads and lost sales which is less than one-for-one.[3] But often these rates are arbitrarily imposed and not based on any real evidence on consumer decision-making.

While substitution rates represent a more honest approach by which to measure potential sales displacement effects, many academics and industry observers have offered an alternative perspective that piracy may actually have beneficial effects on sales. For example, if illegal downloading acts as a sample which precedes legal paid consumption, or if there are bandwagon effects in demand from shared word-of-mouth. In fact, relatively simple economic theory can predict either a positive or negative legal consumption effect from piracy and which effect dominates is ultimately an empirical question.[4] As

[1]Copy Culture in the US and Germany, The American Assembly, 2011.

[2]To put these figures in perspective, total Australian box office revenue in 2011 was A$1.09b, a drop of 3% over 2010's record of A$1.13b (Motion Pictures Distributors Association of Australia, 2012).

[3]See, for example, TERA Consulting (2010) report on digital piracy's impacts in the EU.

[4]Peitz and Waelbroeck (2006) provide a useful survey on the theoretical contributions related to digital

Dejean (2009) and Waldfogel (2012) discuss in detail, three broad approaches have been pursued in the empirical literature investigating the effects of digital piracy on legal paid consumption. First, a number of studies have examined aggregate sales vis-à-vis internet usage, or computer ownership, as a proxy for downloading activity. These studies typically pursue either a cross-sectional approach (e.g. Peitz and Waelbroeck, 2004; Zentner, 2005; Walls, 2008), a time-series approach (e.g. Stevans and Sessions, 2005), or a combination thereof (e.g. Michel, 2006; Liebowitz, 2008). Second, some studies have utilised actual download information (e.g. Liebowtiz, 2006; Bhattacharjee *et al*, 2007; Oberholzer-Gee and Strumpf, 2007; De Vany and Walls, 2007). Third, others have based analyses on data obtained by surveying individuals on their consumption behaviour (e.g. Zentner, 2006; Rob and Waldfogel, 2006, 2007; Hennig-Thurau, Henning and Sattler, 2007; Waldfogel, 2010).

Conceptually, the best approach would appear to be the second one, where actual downloading activity is measured directly and related to legitimate sales. However, this approach is complicated by the inherent simultaneity between sales and downloads. A particularly impressive study employing data on actual downloads vis-à-vis sales is that of Oberholzer-Gee and Strumpf (2007). Their study of the US recorded music industry used file-sharing data collected from OpenNap, a centralised peer-to-peer (P2P) network, providing a sample capturing 0.01% of the world's downloads, and contemporaneous album sales (retail and on-line) over a 17 week period in late 2002. To mitigate the endogeneity between sales and downloads, they considered the number of German school-aged children on holiday under the assumption that German students provided much of the supply of songs on file-sharing networks. However, they were unable to detect any displacement effect between download activity and music sales concluding that the observed decline in music sales is not the primary result of file-sharing.[5]

Our study similarly investigates digital piracy using actual download and sales data, but with specific application to the theatrical film industry. We employ an extensive data set of daily Australian state/territory level P2P torrent downloads and contemporaneous box office revenues. Our study is the first that we know of to consider digital piracy in the film industry using a large data set of actual downloading activity. Our empirical methodology is in many respects similar to the approach of Oberholzer-Gee and Strumpf. However, as discussed further below, there are a number of subtle and important differences. Similar to Olberholzer-Gee and Strumpf, our identification strategy uses the notion of a download cost function which has variation both in film and time dimension, and also allows for structural changes observed in the Australian internet market over the sample period which have likely reduced costs associated with downloading. At the individual film level, we consider downloading cost to be inversely related to the number of peers in the P2P swarm and positively related to download file size. We proxy the number of peers in the download swarm by using the contemporaneous number of downloads occurring in geographically separated Australian states/territories—an approach that has some simi-

piracy.

[5]Although appearing as the lead article in the *Journal of Political Economy*, and having been one of the most downloaded and cited papers in the journal since its publication, the findings have not been accepted without criticism. Notably, Liebowitz (2007, 2010) questions the data construction and instruments used in the econometric component of the research.

larity to the use of (average) price in other markets sometimes employed in differentiated goods models of demand (e.g Hausman, 1994; Nevo, 2001). As discussed in further detail below, this identification strategy is greatly assisted by the panel structure of our data which allows us to account for film-level heterogeneity with film fixed effects. At a more temporal level, the large increases broadband subscriptions, speeds and capacities offered by Australian ISPs allows us to introduce an additional moment condition into our model.

Like Oberholzer-Gee and Strumpf, we find no evidence of a *contemporaneous* relationship between downloading and sales (i.e. box office revenues), but we do find evidence of a sales displacement effect when downloads are considered as a *dynamic stock* over one, two, three and four week windows. We also observe that both contemporaneous and dynamic stock downloads have a significant negative impact on first week box office. Given many films are subject to a release lag between the US and Australian markets, this suggests downloading activity post-US release but pre-Australian release decreases opening week revenues which are well-known to be particularly important in a film's life.[6] We find that the release delay between the US and Australian markets provides an opportunistic window for online pirates which is statistically related to decreased revenues at the box office. Although the present impact on box office revenues appears small, with the increasing use of file-sharing technology and increased speed of bandwith in Australia, this problem is likely to increase. The trend towards day-and-date releases seems the most sensible response given this increasing threat in the absence of legal solutions.

2 Australian Context

Australia provides an interesting context within which to study digital piracy—and in particular that related to theatrical films. Australians are well-known to be some of the most frequent cinema-goers worldwide. According to statistics compiled by Screen Australia, in 2010 Australia ranked third behind Iceland and Singapore in terms of annual admissions per capita, with the US ranking fourth.[7] However, it is also widely known that Australians are among some of the most avid users of P2P file-sharing technologies for music, television and film.[8] Australia's attraction to file-sharing is often attributed to relatively high content prices as well as international release delays for TV and film.[9] With the National Broadband Network (NBN) progressively rolled-out over the next decade, content industries fear that digital piracy will proliferate even further as consumers are able to access and download illegal content with increasing speed and ease.

Although Australian content providers have been lobbying the government to take a stand against piracy, thus far their efforts have largely been ignored with policy makers

[6]This evidence is consistent with a recent study by Danaher and Waldfogel (2012) who find that, on average, international box office revenues are 7% lower when international releases are delayed relative to the US market in a study of seventeen countries pre and post BitTorrent technology.

[7]See `http://www.screenaustralia.com.au/research/statistics/acompadmitper.asp`.

[8]For example, the HBO hit *Game of Thrones* was most heavily downloaded by Australians (Ernesto, 2012). In relation to music downloads, a recent study by MusicMetric found that Australians download more songs per-capita than any other country and ranked sixth overall in terms of volume (Zuel, 2012).

[9]The disparity in prices for digital content has recently caught the attention of policy makers who have commenced an inquiry within the Australian House of Representatives about the issue.

encouraging continued negotiations with ISPs to find a cooperative solution.[10] As a result of political inaction, some content owners have pursued direct legal action against individuals and, subsequently, ISPs in their war on piracy. The initial legal strategy pursued by some companies involved actions previously employed in the US (and other countries) by sending Australian ISPs 'cease and desist' notices outlining details of the customer's infringement and requesting they threaten the individual with disconnection of their internet service.[11] However, not all ISPs complied with such instruction and the failure to comply by Australia's second largest ISP, iiNet, resulted in a landmark court case spanning more than four years.

In November 2008, a consortium of 34 record labels, pay-TV providers, film studios and other content providers filed the case against iiNet for failing to discipline its customers in relation to allegations of copyright infringement. The case of *Roadshow Films and others v iiNet* (or commonly *AFACT v iiNet*) was initially heard by the Federal Court of Australia and decided on 4 February 2010 with the trial judge ruling in favour of iiNet and awarding costs. In passing judgement, the trial judge noted that while iiNet users did infringe copyright, it was not the responsibility of iiNet to police its customers on the infringement of other parties' copyrights. The decision was subsequently appealed by AFACT to the full bench (Full Court) of the Federal Court on 24 February 2011 but was again dismissed by the presiding judges. The trial judges upheld the initial decision but for different reasons. They noted that although iiNet showed an indifferent attitude to the complainants' allegations, iiNet's inaction did not constitute authorisation for the act of copyright infringement. On further appeal, the case was heard by the High (Supreme) Court of Australia which again sided with iiNet in judgement passed on 20 April 2012. The Court unanimously dismissed AFACT's appeal and ordered AFACT to pay costs of approximately A$9m.[12] Given the Courts' decisions on this case, we believe it is unlikely downloading behaviour would have reduced in response to these rulings.

Although it is undeniably true that Australians are among the world's most prolific users of P2P file-sharing technologies, it is also true that until recently—and certainly over the sample period of our study—many internet users faced constraints on the activity given typical internet plans contained data download limits which were relatively low in comparison to other countries.[13] Given a typical film download can be in excess of one gigabyte, consumers with limited data allowances would have to be considerate of the amount of downloading performed each month so as not to incur excess data charges or have download speed throttled.[14] It would therefore seem reasonable to believe that such

[10] The 'three-strikes' policy adopted in a number of countries (e.g. UK and France) was considered in Australia; however, to date, such a policy has found little support from policy makers.

[11] For example, the US television network CBS sent infringement notices to one Australian ISP, TPG, in relation to alleged offences by their subscribers warning them "1) Remove or disable access to the individual who has engaged in the conduct; and 2) Take appropriate action against the individual under your Abuse of Policy Terms of Service Agreement" (Britton, 2011).

[12] See `http://www.iinet.net.au/about/mediacentre/releases/index.html`

[13] While Australian internet plans vary considerably with price, most entry level plans provided 10Gb per month or less over the sample period of our study.

[14] Different Australian ISPs have differing policies in relation to exceeding download limits. While some charge excess data at a premium, others restrict download speeds to those of a dial-up connection (e.g. 256Kbps).

constraints would incline internet users towards downloading more highly valued content, such as new films playing at cinemas.

However, while many internet users have been constrained by download limits in Australia, the particular period of our study also coincides with dramatic increases in data allowances offered by all the major Australian ISPs.[15] Data collected by the Australian Bureau of Statistics (ABS) on internet subscriptions, speeds and downloads supports large increases in all these areas for all Australian ISPs operating during our sample period. In relation to broadband subscriptions, from December 2009 to June 2011 there was an increase from 8.0m to 10.3m (i.e. 28%) subscribers and an increase from 2.5m to 5.0m (i.e. 102%) subscribers for plans offering above >8Mb/second. In relation to actual volume of data downloaded, the quarterly volume increased from 127,661Tb (December 2009 quarter) to 274,096Tb (June 2011 quarter)—an increase of more than 114%.[16] These aggregate statistics provide evidence of dramatically changing landscape in internet service provision in Australia as the overall number of broadband subscriptions increased coupled with faster download speeds and less restrictive data allowances. *Ceteris paribus*, these changes would almost certainly increase levels of downloading observed over our sample period. We make use of this exogenous structural change in the empirical model we outline below.

3 Data and Descriptive Statistics

To investigate the impact of file-sharing on film revenues, we employ an extensive data set of Australian state/territory level daily box office revenues and P2P torrent downloads of 166 films released in Australian cinemas between January 2010 and August 2011. The films in our sample are typically large budget 'Hollywood-type' films which received a wide-release in the US theatrical market as well as an Australian theatrical release. Given the international nature of these films, *a-priori* we would expect substantial interest in both cinematic consumption and illegal downloading allowing us to investigate potential displacement relationships between downloading activity and box office revenues. The torrent data were sourced from Peer Media Technology—a company which, among other services, measures digital piracy for companies in the entertainment, software and publishing industries.[17] We tracked downloads on three popular P2P networks: 1) BitTorrent, 2) eDonkey, and 3) Ares, where a download is defined as a unique instance of an IP address attempting to download an appropriately named file on a given day. The IP addresses are subsequently geo-located by another company, MaxMind, to provide state/territory-level number of downloads per title per day in our particular context. Peer Media Technology estimates that their measurement provides approximately 55% of all downloads in the

[15] For example, at January 1, 2010, Australia's third largest ISP Optus offered broadband internet plans ranging in price from A$85/month to A$130/month which provided from 4Gb to 30Gb download allowance per month. For the same prices, data was increased to a range of 7Gb-50Gb in April 2010 and further increased to 120Gb to 1Tb in December 2010.

[16] See ABS Catalogue 8153.0

[17] Peer Media Technology ceased operations in December 2012.

Australian context.[18]

The torrent data of each film in our study span a longer period than the observed Australian theatrical life of the film, allowing us to track downloads which may have occurred both before and after the theatrical window. In particular, we observe downloads post US release, but pre-Australian release. Figure 1 reveals that the number of downloads spikes after the initial US release, presumably the result of increased availability and interest. But a significant number of downloads occurs prior to the Australian release and we hypothesize that this is related to release gaps between the two markets—an issue we shall discuss in more detail below. During the theatrical release window, we observe contemporaneous (daily) box office revenues and number of theatres for each film disaggregated to the state/territory level (Rentrak). In addition, we also observe US box office revenues, opening theatres and cinematic release dates, hence we observe the release gap between the US and Australian theatrical releases.[19]

In total we observe 295,304 torrent download data points and 64,328 daily box office revenue and theatre data points. We limit our attention to 20 weeks of box office revenues post Australian release which provides us 56,663 data points in the final estimation sample. Tables 1 and 2 provide summary statistics for our data. Table 1 details aggregate level information for the 166 films observed. On average, each film earned nearly A$8.9m at the Australian box office and was released on 259 screens. The highest earning film, *Harry Potter and the Deathly Hallows Part 2*, made almost A$51m. The average number of downloads per title was almost 113,000, with an average file-size of 1.2Gb. The most downloaded film, *Inception*, was downloaded more than 435,000 times. As noted above, all the films in our sample had a wide-release in the US market of at least 2,000 theatres (average of 3,125), and were generally large budget titles with an average (estimated) production budget of US$70m (data sourced from IMDb). Of all 166 films we observe, the correlation between total revenue and total downloads is moderately strong at 0.522. Figure 2 shows this relation with a simple linear regression which reveals a statistically significant positive relation.

Table 2 provides summary statistics for the disaggregated data used in estimation. Of the 56,663 data points, the average film's daily (state/territory level) revenue is A$25,148 and the average number of downloads is 95. These are simply weighted averages of the state/territory data contained in the body of the table. The state/territory summary data is consistent with respective sate/territory populations in terms of mean ranking.[20] However, the data for Northern Territory (NT) downloads are particularly low due to problems in primary data collection.[21] Aggregating all revenue and downloads across states/territories for all films observed on each of the 596 days in our sample, the average Australia-wide daily revenue was just under A$2.5m, with the highest recorded single day revenue of A$9.4m occurring on Wednesday July 13, 2011, coinciding with the release

[18] In our empirical model, we implicitly assume that there is no systematic relation between the choice of P2P network and patterns of demand.

[19] All of the theatrical market data on films were sourced from Rentrak.

[20] In December 2011, New South Wales (NSW) and Victoria (VIC) were the largest states with populations of 7.25 million (approximately 32% of Australia's population) and 5.57 million (approximately 25%), respectively. Tasmania (TAS), Australian Capital Territory (ACT), and Northern Territory (NT) are the smallest in terms of population with 0.51 million, 0.37 million and 0.23 million, respectively.

[21] All estimates were re-calculated omitting this territory; there were no qualitative changes to results.

of *Harry Potter and the Deathly Hallows*. In terms of downloads, nation-wide the daily average was just over 31,000 with the highest single day number of downloads occurring on Sunday June 12, 2011, where more than 61,700 downloads occurred. Our data also displays intra-week seasonality in relation to both revenue and downloads. Unsurprisingly Saturday, Sundays and Friday recorded highest average nation-wide revenues (downloads) at A$4.0m (35,296), A$3.3m (34,863), and A$2.7 (28,424), respectively. We control for the intra-week seasonality in our model with the use of day-of-week dummy variables as discussed in the following section.[22]

4 Econometric Model

Contemporaneous Downloads

Our empirical model aims to quantify the sales displacement effect from illegal P2P torrent downloads on box office revenue. The approach is similar to Oberholzer-Gee and Strumpf (2007), with a number of important differences. First, and most obviously, our context is theatrical film revenues rather than music sales. This removes the complicated issue of transforming single downloads (sales) to a proxy for album downloads (sales). Second, the data are observed at daily, rather than weekly, levels. Also, our data are observed at the state/territory level, rather than national level. Third, our identification strategy relies on a downloading cost function which relates to contemporaneous downloads in other Australian states (proxying for number of peers), download file-size and significant structural changes in the Australian internet service industry over our sample period. We discuss identification in more detail below.

Our empirical model is based on the following equation

$$\ln R_{ist} = \varphi \ln D_{ist} + \mathbf{x'}_{ist}\beta + \mathbf{u'}_{ist}\delta + \eta_s + v_i + \varepsilon_{ist} \tag{1}$$

where R_{ist} and D_{ist} define revenue and downloads of film i in state s on date t, respectively. The logarithmic transformation is applied to revenue and download data since both distributions are bounded below at zero and are right skewed. An additional benefit of the log-transformation is that the impact of downloads on revenues can be interpreted as an elasticity. We partition control variables into observable and unobservable vectors $\mathbf{x'}_{ist}$ and $\mathbf{u'}_{ist}$, respectively. In particular, vector $\mathbf{x'}_{ist}$ includes the (time-variant) number of theatres showing film i in state s on date t (TH_{ist}) which also enters in log-form; the week of the run of film i in state s on date t (WK_{ist}); as well as a set of dummy variables for day-of-week effects (DW^d, where d indexes day). Additionally, we include state/territory fixed-effects η_s and film fixed-effects v_i.[23]

We seek to make causal inference on the parameter φ. However, because popular films at the box office are also likely to be popular on P2P networks, download activity should

[22] Day-of-week dummy variables control for intra-week patterns in demand which typically reveal spikes on weekends and Tuesdays, which is a discount day for most cinemas in Australia (see De Roos and McKenzie, 2014).

[23] As discussed below, the inclusion of film fixed-effects goes a long way to purging the simultaneity between revenue and downloads.

be treated as endogenous. In terms of (1), this implies that $E\left(\mathbf{u'}_{ist}\delta + \varepsilon_{ist}|D_{ist}\right) \neq 0$. To remedy this problem, we employ two-step efficient generalized method of moments (GMM) estimation with robust standard errors (Hansen, 1982).[24]

Our instruments are chosen based on the assumption of a downloading cost function with 'cost-shifters' which would not be expected to be related to demand for theatrical consumption. We propose the following

$$C^D_{ist} = f(P_{it}, FS_i, t) \tag{2}$$

where C^D_{ist}, P_{ist} and FS_i are cost of download, (number of) peers, and file size, respectively. Again, t represents time index. The first argument of the cost function relates to the number of peers who are (contemporaneously) 'seeding' the film file. Typically, the more seeders in a swarm, the faster the download and the more likely the chance of a successful download. Therefore, we would expect $\partial C/\partial P < 0$. Although we don't observe actual information on the number of peers, we proxy this variable with the (contemporaneous) total number of downloads made across all other Australian states on date t. Regarding the second argument, we expect downloading costs increase as the size of the (movie) file being downloaded increases, or $\partial C/\partial FS > 0$. This is assuming additional time and opportunity costs (for data constrained users) for larger downloads. The final argument is a simple time trend. As reported in Section 2, the period 2010-11 saw significant increases in broadband subscriptions, speeds and data allowances in Australia. As a result, we believe downloading costs have fallen over our sample and conjecture $\partial C/\partial t < 0$. We discuss identification issues in more detail in the following section.

Identification

Identification in our model derives from the inclusion of instruments in the GMM estimation procedure which relate to downloading activity in other states (as a proxy for total number of peers), file size of film file to be downloaded, and a trend variable to account for structural changes in the Australian internet industry.

There are three well-known requirements for a good instrument: 1) that it itself does not belong in the focal relationship, 2) that it is uncorrelated with the error process, and 3) that it is correlated with the endogenous variable. Taken together, the instrument's

[24] Define the general form of equation (1) as $\mathrm{y} = \mathrm{X}\beta + \varepsilon$, where the matrix of regressors, X, is of dimensions $N \times k$, and where N is the number of observations and k is the number of regressors. Also define the covariance matrix of ε as $E[\varepsilon\varepsilon'|\mathrm{X}] = \Omega$, which is of dimension $N \times N$. As some regressors (*i.e.* downloads) are endogenous, partition the regressors $\{\mathrm{x}_1\mathrm{x}_2\}$ with the k_1 regressors x_1 endogenous and the $(k-k_1)$ remaining regressors x_2 assumed to be exogenous. The matrix of instrumental variables Z is $N \times l$, where the instruments are partitioned into $\{\mathrm{z}_1\mathrm{z}_2\}$, where the l_1 instruments z_1 are excluded instruments and the remaining $(l - l_1)$ instrument $\mathrm{z}_2 \equiv \mathrm{x}_2$ are the included instruments (or exogenous regressors). The l instruments give a set of moment conditions, $E[\mathrm{z}\varepsilon] = 0$, with the sample analogue $\overline{g}(\beta) = (1/N)\mathrm{Z}'\varepsilon$. The intuition is to select a estimator for β which solves $\overline{g}(\hat{\beta}_{GMM}) = 0$. The GMM estimator chooses $\hat{\beta}$ that minimises the GMM objective function $J(\hat{\beta}) = N\overline{g}(\hat{\beta})'\mathrm{W}\overline{g}(\hat{\beta})$, where W is an $l \times l$ weighting matrix. The optimal weighting matrix is that which produces the most efficient estimates. Following Hansen (1982), $\mathrm{W} = \mathrm{S}^{-1}$, where S is the covariance matrix of the moment conditions $\mathrm{S} = E[\mathrm{Z}'\Omega\mathrm{Z}]$. For the heteroskedastic-consistent estimate of S, $\hat{\mathrm{S}} = (1/N)\Sigma_{i=1}^{N}\hat{\varepsilon}_i^2\mathrm{Z}_i'\mathrm{Z}_i$ is used. The residuals come from a 2SLS regression which provides consistent estimates of β.

effect on the outcome variable occurs only through its association with the endogenous variable. Our first two instruments clearly satisfy the first requirement. By assumption of geographic separation, downloads in other states couldn't displace sales in a focal state. And file-size certainly has no place in a model of box office revenues. As for trend, we note three reasons why we do not believe trend should appear in the focal (revenue) relation: 1) the relatively short time frame of our sample (20 months), 2) no significant increase in cinema infrastructure, and 3) slightly declining total sales as measured by a total decline in box office of 3% from 2010 to 2011 (MPDAA). With the significant changes in internet provision previously discussed, it is appropriate to consider the trend as also satisfying the first requirement outlined above.

The second and third requirements deserve further attention. Regarding the first instrument, the use of summed downloads in other Australian states is taken as a proxy for the number of peers (seeds) P2P network and, as such, should be inversely related to downloading costs. That is, more downloads in other states suggests more seeders generally on a network and as a result a higher chance of a quicker and successful download. The use of the total number of downloads from geographically separated markets as an instrument is somewhat similar to the use of (average) price from other markets when modelling demand for differentiated goods (e.g. Hausman, 1994; Nevo, 2001). The intuition is that prices are correlated through common marginal cost shocks. Assuming the errors in demand are independent across market, this 'cost shifter' approach is valid.

In our context, we similarly require an independence assumption for total downloads in other states to be a valid instrument. Specifically, the unobserved elements of $\mathbf{u}'_{ist}$ in (1) should not also be correlated with our instrument. We argue this is true based largely on the panel structure of our data permitting us to use film fixed-effects to absorb much of the film level heterogeneity, which is the major source of our endogeneity.[25] We therefore implicitly assume that the remaining (time) variation in the instrument is then more reflective of technical aspects and shocks related to downloading than any systematic variable which relates to unobserved demand.[26] For example, a low number of downloads could reflect a poor quality torrent file (e.g. camcorder version) or a problematic tracker (i.e. the server which directs the torrent uploads and downloads not responding).[27] As a result, low number of peers may indicate high chances of download failure and associated higher cost (wasted time and data capacity). It is also typically true that more seeders implies a faster download with higher probability of success. Downloaders therefore preference torrents for which they can see a high seeder ratios. At the margin therefore downloaders would prefer to download titles with more active swarms.

Regarding the second instrument, file size should have a direct impact on downloading activity particularly for internet users who are constrained by data limits. We conjecture the marginal consumer is less likely to download a large file if it is going to compromise

[25] As the results below reveal, we establish a sales displacement effect in our OLS models simply with the inclusion of film fixed effects which, with the inclusion of week of theatrical run, resolve much of the simultaneity problem we face.

[26] In essence, we are assuming that there don't exist any unobserved temporal film effects which affect both downloads and revenues simultaneously. Or alternatively, the (unobserved) quality of a film remains constant through time albeit with declining sales as the potential market becomes saturated. This feature of demand is captured in the model by the week-of-run variable.

[27] We are implicitly assuming users become aware of this through message boards or some other signal.

their download limit. Additionally, users may be dissuaded from downloading large files due to the time involved.

The third instrument we make use of is related to the changing structural features of the Australian internet landscape. As discussed in Section 2, there were significant increases in data offerings and speeds from Australian ISPs over the sample period. As a direct result, broadband subscriptions, speeds and data downloaded jumped dramatically. In this environment, it would seem that illegal downloading activity would also have increased. To capture this we include a simple linear trend as an additional instrument in our model.

Dynamic Stock of Downloads

The model described in (1) captures the potential sales displacement for film i, in market s, on date t. There are, however, a number of reasons it would seem more appropriate to consider the total number of downloads over some window of time prior to, and including, date t rather than only those downloads observed contemporaneously. For example, illegal downloaders may browse torrent sites for new titles and may initialise a download with the intent of viewing the download at a later time—and, we assume, forego the cinematic alternative. In addition, downloading speeds (at least during the period of analysis in Australia) could be restrictively slow preventing instant playback necessitating files are downloaded in advance of actual consumption. Further, there is also the possibility that once a download has been completed, an individual may share the file with friends which would similarly contribute to potential future sales displacement effects. For these reason, we consider it likely that downloading over some window of time affects future box office sales which would not be captured in our contemporaneous specification. We subsequently refer to this alternate approach as a *dynamic stock of downloads* in our empirical methodology and consider the modified model:

$$\ln R_{ist} = \varphi \ln D_{isT} + \mathbf{x'}_{ist}\beta + \mathbf{u'}_{ist}\delta + \eta_s + v_i + \varepsilon_{ist} \tag{3}$$

where D_{isT} defines downloads of film i in state s over a period of T days, where $T \in \{7, 14, 21, 28\}$, prior to and including the date of observation. Alternatively stated, $D_{isT} = \sum_{r=0}^{T} D_{is(t-r)}$. Essentially, the modification simply considers the dynamic stock of downloads over one, two, three or four weeks prior to (and including) date t. Our first instrument related to number of downloaders in other states, as discussed above, $D_{is'T}$, is also similarly defined and now represents total downloads of film i in all other states/territories $s' \neq s$ over T, or $D_{is'T} = \sum_{r=0}^{T} D_{is'(t-r)}$. All other variables remain as specified above.

This transformation requires further discussion in relation to role of our instruments. In particular, how to interpret the the dynamic stock variable in the cost function. The number of downloads per T are now more reflective of aggregated popularity of the torrent on the P2P networks. The time-invariant heterogeneity will, once again, be picked up though film fixed-effect. But the temporal variation will pick up changes in technical aspects of the torrent file. As an example, take the case of a typical film at the box office which is enjoying an average reception and theatre/screens decease familiarly over some number of weeks. Suppose up until week three only a bad (camcorder) torrent was available for this film. As a result, educated downloaders avoided this torrent and only low downloads were observed. Beyond the third week, however, suppose a better torrent appeared which led to higher numbers of downloads. This temporal variation is entirely related to the changes in the P2P offerings. The primary instrument is perfectly constructed to capture this type of variation. We continue discussion of identification in Section 6.

5 Estimation Results

Contemporaneous Downloads

Table 3 provides GMM estimation results for the base model where downloads are treated contemporaneously to revenue (i.e. on the same date of observation). The first column reports simple OLS estimates without film fixed-effects. The coefficient on the key variable of interest, contemporaneous downloads (D_{ist}), is positive and significantly different from zero at 1%. The other estimated coefficients conform with a-priori expectations. Specifically, week-of-run (WK_{ist}) and contemporaneous theatres (TH_{ist}) reveal negative and positive signage, respectively. Again, both are statistically significant. Inclusion of film fixed-effects in the second column of Table 3 purges some of the endogeneity between revenue and downloads that exists because of unobservable shared tastes for more popular films. Notably, the estimated coefficient on contemporaneous downloads is reduced but is still significantly positive. The third and fourth columns represent the first and second stage GMM estimation defined in equation (1), respectively. The first stage estimates reveal a strong and significant positive relation with $D_{is't}$, but insignificant relation with file size and trend (although signage is as expected). Using the Kleibergen-Paap (2006) LM test for under-identification, Cragg and Donald (1993) test for weak identification, and the Hansen J test for over-identification (see, for example, Hayashi, 2000) we find our instruments reject under- and weak-identification but do not reject over-identification. However, it is well known that this test suffers in large samples.

The second stage results of the GMM estimation reveals a further decrease in the magnitude of the estimated coefficient on contemporaneous downloads when compared to the OLS specification with film fixed effects. The positive relation between downloading activity and revenues is still apparent and significant, but the magnitude of the effect has been substantially reduced. This supports that the excluded variables (instruments) are helping to identify the model by reducing the positive bias induced by the endogeneity of downloads. The other estimated coefficients in the second stage regressions are similar in magnitude and signage to the OLS results with film fixed-effects of column 2.

Time-invariant and film-specific variables—such as budget, cast/director appeal, (pre-release) advertising, genre, classification rating, etc—are implicitly captured in our model by the inclusion of film fixed-effects. To examine the contributions of some of these variables, we extract individual film fixed-effects and consider them against the time-invariant film-specific variables observed in our data set. The correlation between budget and the extracted fixed-effects is 0.50 (compared to a correlation of 0.65 with total film revenues); and the correlation between (national) opening week screens and fixed-effects is 0.72 (compared to a correlation of 0.82 with total film revenues). In a basic OLS regression with fixed-effects as the dependent variable, estimated coefficients of both (log) budget and (log) opening screens are positive and significant at the 1% level of significance with an R^2=0.44. Both relationships still hold at 1% when controls are added for the categorical variables of sequel, genre, and classification rating with R^2=0.64. In comparison, a regression of (log) total revenue on the same set of covariates found similar explanatory evidence in terms of signage and significance with R^2=0.81. To the extent that the extracted fixed-effects correlate strongly with key variables, which have been shown as important attributes of overall film revenues, these findings validate the use of film fixed-effects as serving the model to capture the time-invariant determinants of demand for the films we observe.[28]

Dynamic Stock of Downloads

As discussed in Section 4, we consider it likely that individuals make theatre attendance decisions after a download decision. Under this assumption, it would be more appropriate to consider the number of downloads over some window of time prior to the actual date at which box office is observed. Table 4 provides evidence when this window of time is considered at one, two, three and four weeks prior to (and including) the date of observation. We report only results from the fixed effects and GMM models with all three instruments (i.e. sum of downloads from other states, file size and trend). In all four models, T=7,14,21,28, the (log) downloads coefficient from the GMM estimation is less than the OLS fixed-effects (as also observed within the contemporaneous results reported in Table 3) providing evidence of a reduction in the inherent bias of this coefficient. However, unlike the contemporaneous case, the relationship between downloads and revenues is now observed to be statistically negative implying a sales displacement effect. The (absolute) increasing magnitude of the download coefficient over the four models, T=7,14,21,28, is a simple manifestation of the increasing dynamic stock. The estimated coefficient suggests a sales displacement elasticity in the range 0.06-0.3. We discuss the economic interpretation of these results further below.

Regarding the first stage results, the primary instrument relating to download activity in other states continues to be positive (as expected) and highly statistically significant across all four models. Also, the file size and trend variables now become highly significant with negative and positive signage, respectively. This is consistent with our *a-priori* intuition related to the cost of downloading being positively related to the size of a download file, and also that download costs have fallen due to the changing structural features in the Australian internet service provision industry.

[28] A growing literature of empirical research has examined the correlates of financially successful films such as budgets, advertising and publicity expenditures, opening screens/theatres, marquee stars, critical reviews, awards, prequels/sequels, genre, rating, etc. See survey of McKenzie (2012) for examples.

6 Discussion of Estimation Results

Instruments and Identification

Correct causal inference depends critically on the model being correctly specified. In modelling a potential sales displacement effect, the inherent difficulty lies in purging the simultaneity between downloads and revenues which manifests in a positive bias on the estimated coefficient in the absence of remedial measures. Fixed-effects provide a partial solution in panel data sets such as ours because the (time invariant) heterogeneity between films can be accommodated. However, full identification of the effect requires an instrumental variable which also has some temporal variation and therefore removes time-varying unobserved heterogeneous effects. We argue that our instruments are particularly strong and satisfy both economic and statistical requirements. Intuitively, the identification in our model stems from an assumption that temporal levels of downloading relate more to technical aspects of file-sharing rather than unobserved temporal shocks to demand. This is not to say that there is no unobserved relation between box office revenues and downloads, but simply to say that it is not one that is systematically time varying and is thus accounted for by film fixed-effects.

It is worth further interrogating this critical assumption. As discussed previously, there is similarity in approach to the use of (average) price in geographically separated markets (aka Hausman-type instruments) to instrument price in a focal market when modelling demand for differentiated products. However, many researchers have also noted this approach becomes invalid if unobserved demand shocks are correlated between markets. For example, national advertising/marketing campaigns may generate an unobserved shock to demand and could potentially increase prices. In our context, advertising is one possible source of time-varying heterogeneity which may affect both downloads and revenues and isn't observed in our model. However, the large majority of typical advertising spends on a movie occur pre-release (typically in excess of 90%) and therefore would not affect demand temporally beyond what is captured by the film fixed-effect. Similar concerns may also be put forward for other types of post-release promotion of the film. This could, for example, be critics' reviews or award nominations and/or wins. While we cannot rule out these effects entirely, most films are reviewed prior to (or coinciding with) opening weekend. Additionally, the release delay between the US and Australian markets means many reviews are already available online prior to the Australian release. Award nominations/wins are potentially an issue that could cause an unobserved demand shock to both revenues and downloads. However, the relatively low number of films in our sample which were playing in cinemas when major award nominations and/or wins took place leads us to believe this would not be a significant effect. Moreover, the types of films we consider in our sampling frame (i.e. US wide release films) are often not those which are typically nominated or receive major awards.[29]

While we have explained the theoretical justification why our instruments—in particular, the summed downloads of other states—are valid, it is also of value to evaluate the statistical evidence a little more closely. The results of the first stage regressions

[29]For example, *The King's Speech* which won Best Picture in the 2011 Oscars was not part of our sample as it only received a limited opening in the US market.

reveal a particularly high R^2 which indicates high explanatory power of the (included and excluded) instruments. This may create concern that the correlation is too high in the sense that the first stage is just recovering the endogenous variable. This would certainly be true if there was near perfect correlation between downloads in the focal market and other states. However, a simple regressions of $\log D_{isT}$ on $\log D_{is'T}$ for T=1,7,14,28 reveal R^2 in the range 0.22-0.34. In terms of simple correlations, the range is 0.47-0.58 between the endogenous (log) download variable and respective (log) downloads in other states. It is apparent that even though there is reasonably strong correlation between the endogenous variable and the instrument—the statistical requirement—it is far from simply recovering itself and the other instruments play an important role. As well, in terms of the statistical requirement that the instrument is uncorrelated with revenues, the condition appears to be well satisfied with simple correlation in the range -0.11 to 0.08. More formally, it was also observed that the statistical tests rejected under-identification and weak-identification, but not over-identification. However, as noted above, we believe this is in part due the large sample size.

Opening Week Revenues and Release Delay

One potential criticism of our model is that we are observing daily film revenues over the theatrical life of a film, which would typically be decreasing, whereas the dynamic stock variable may be increasing. Although we include a week-of-run variable in both the first and second stage regressions, this potentially inverse relation may be driving the results. To address this potential issue, we restrict the model described in equations (1 to only model revenues of films in their opening week of theatrical release. This means the variable WK_{ist} is now redundant as all films are observed in their first week of release. Table 5 provides results for the second stage regressions for the contemporaneous and dynamic models. In all cases the coefficient on the downloads variable is statistically less than zero implying that first week revenues are subject to a sales displacement effect. As illustrated graphically in Figure 1, the fact that many films are subject to a release delay between the US and Australian markets seems a logical explanation of the decreased first week sales when there is an opportunity for download prior to the opportunity for legal consumption.[30] In our sample of 166 films, the average release gap between the US and Australian markets was 28 days (median of 13 days). One film, *Thor*, was released in Australia two weeks prior to the US release, a further six films were released in Australian cinemas one week prior to the US release, and 50 films had a simultaneous release with the US release (opening within one day of the US release). The remainder of films had a positive release gap with the greatest gap being *Diary of a Wimpy Kid* which opened in Australia cinemas six months after its US release.

A simple regression of (log) downloads (observed within the first week) on observation date relative to the US release (controlling for day-of-week, state and film), retrieved a significant positive relation suggesting an increase in the number of first week downloads

[30] Even in the absence of a release gap, torrent sites sometimes feature leaked studio copies of titles which may have been intended for promotional reasons or award consideration. This means that it is possible for illegal downloading to occur prior to official international (as well as domestic Australian) release.

the more time elapsed since the US release, *ceteris paribus*. Given that illegal copies of films typically show up after the US release, this evidence is supportive of the release delay between the US and Australian markets providing opportunities for illegal consumption before the film is theatrically released in Australia. When release delays between the US and Australian markets are significant, the likelihood of high quality torrents files appearing on networks increases and increased levels of downloading would displace more sales.

Weekly Model

The model outlined in Section 4 was estimated with daily data. To examine whether this feature of our data has any bearing on results we also consider a model where revenue and downloads are considered at the week level, rather than daily. We do this both for the contemporaneous and dynamic stock models. The contemporaneous model is analogous to (1) but now the time subscript t represents a week, rather than a day, and is redefined t^W. The theatres variable, TH_{ist}, is redefined so it now represent the maximum number of theatres screening on any day of that week for the film of interest. Also, the day-of-week dummy variables are now redundant. The dynamic stock model is similarly redefined in terms of weeks as T^W, where T^W now represents the number of weeks included prior to week t^W. We consider the dynamic horizons T^W=1,2,3,4 implying that downloads are observed for one, two, three and four weeks prior to, and including, week t^W.

Results for the weekly model are reported in Table 6 which display second stage regression results for the contemporaneous and dynamic stock models. In the contemporaneous and T^W=1 model, a positive and significant relation between downloads is observed. However, in the T^W=2,3,4 models the relationship is negative implying a statistically significant sales displacement relationship. The magnitude of the estimated coefficients suggest a displacement elasticity of 0.15% for a 1% increase in downloads over the four weeks prior to, and also including, the week of observation.

Forward Looking Dynamic Stock of Downloads

It has been argued that the dynamic stock approach to considering future sales displacement is more realistic than a contemporaneous approach. It might also be possible, however, that an individual forgoes cinema attendance today by making a decision to download the title at some point in the future. Obviously, this argument relies on the title being available for download and that the intention is actually carried out. To test whether downloaders forgo theatrical consumption prior to actually downloading the file, we consider a simple modification of our previously specified dynamic model (3) in which T is now forward looking. We denote this forward looking dynamic stock as T^F and define $D_{isT^F} = \sum_{r=0}^{T} D_{is(t+r)}$ for the horizons T^F=7,14,21,28. Similarly our instrument is also redefined as $D_{is'T^F} = \sum_{r=0}^{T} D_{is'(t+r)}$ for all $s' \neq s$.

As Table 7 demonstrates, the estimated coefficients on the (log) downloads variable across all specifications (T^F=1,7,14,21,28) remains positive implying no displacement effect. Taken with the evidence of the (backward looking) dynamic stock model, this might suggest individuals who partake in downloading (and substitute it for paid cinematic consumption) are time impatient and seek out films on torrent sites early in their theatrical life—often prior to the official release if the film has already been released overseas as discussed above.[31] This observation is also consistent with the 'movie-maven' subculture among (particularly young and tech savvy) consumers who desire to see a film early and before the masses.

Sales Displacement Effects

Although we have detected a statistically significant negative impact from file-sharing on box office sales, the economic significance of these effects appear relatively small. We demonstrate this by providing some 'back-of-the-envelope' numbers concerning the potential sales displacement effects of piracy implied by our results. Table 8 provides estimates of sales displacement or 'substitution rates' for the daily, opening week, and weekly models as presented in Tables 4, 5, and 6.[32].

If we focus on the daily T=7,14,21,28 models, we observe sales displacement elasticities of 0.06, 0.17, 0.25 and 0.31, respectively. Given the median daily box office of films in our sample is A\$3,593 (Table 2), this translates to a reduction in revenue of between A\$2.23 and A\$11.07, or between 0.17 and 0.86 people assuming an average ticket price of A\$12.87 (Screen Australia reported average ticket price for 2011), for a 1% increase in downloading activity across these time horizons. Given that state/territory level median number of downloads over one, two, three and four weeks are 680, 1345, 1962 and 2485 (adjusting for Peer Media Technology's estimate of a 55% market coverage), this suggests that somewhere between 29 and 40 downloads displaces one purchased ticket each day depending upon which model is considered. Put another way, for every 100 downloads somewhere between 2.5 and 3.5 cinema admissions are displaced.

When considering the opening week and weekly models, we find lower levels of displacement for the weekly model but also find substantially higher levels of sales displacement for opening week revenues. Given the range of elasticities from Table 5, and the median opening week daily revenue of A\$31,046, we estimate a 1% increase in downloads displaces between one and three paid admissions. With the range of the median number of downloads between 371 (T=7) and 725 (T=28), these levels imply that anywhere from 2.5 to 5.9 downloads displaces one paid admission—which appears relatively high—but in terms of economic significance, the overall potential effect is low because of the relatively low levels of downloading actually taking place in the first week. In part this is reflective of the large number of films with simultaneous releases in the US and Australian market, but it is apparent that the longer a film's release is delayed between the two markets, the more likely the title is to appear on torrent sites which in turn increases the number of illegal downloads.

In terms of the weekly model's results of Table 6, and given the median weekly rev-

[31] In some sense the industry's practice of pre-release TV advertising may be contributing to this problem if consumer interest is aroused by trailers and the torrent file has already appeared on P2P networks.

[32] 'Substitution rate' is industry terminology in studies which aim to measure sales displacement effects assuming a less than one for one relation between downloads and lost sales. See, for example, Mitra-Kahn (2011)

enue of A$19,142, the estimated displacement for T=2,3 and 4 (weeks) suggest between 11.7 and 15.9 downloads displaces one paid admission over a weekly time horizon. This finding is lower than the daily model where it took between 29 and 40 downloads over the various windows considered to displace a sale on a given day. In the weekly model, we are observing approximately the same time frame for downloads but now a week for revenue/admissions. However, the calculation is not simply divided by seven as the medians are not related by this scale. As it turns out, the displacement effect is estimated larger in the daily model if the substitution rate is simply multiplied by seven.

To put these 'back of the envelope' calculations in perspective, assume that (as some industry reports implicitly suggest) one download displaces one paid admission. Then the total lost revenue of the median film in is in the order of A$1.3m (assuming median downloads of 100,000—see Table 1—and ticket price of $12.87), or about 17.5%. If we compare to the maximum (daily) substitution rate of 3.5, and continuing to assume median downloads of 100,000 and an average ticket price of A$12.87, this would imply total lost revenues of A$45k (less than 1%) for the median film.

7 Conclusion

This study has investigated digital piracy in the context of the Australian theatrical film industry. We find evidence of a sales displacement effect from illegal downloading on box office revenues. However, our estimates suggest (at least at present) the economic magnitude of this effect is small. One particular issue our study sheds light on is that piracy behaviour increases proportionally to the release gap between the US and Australian markets. Opening week revenues were shown to decline significantly because of downloads which occurred prior to the theatrical release. This finding is not unsurprising and provides partial explanation for the observed and growing trend of day-and-date world-wide releases—particularly for blockbuster titles.

Whether the theatrical film industry is likely to suffer revenue declines similar to those observed in the music industry is yet to be seen. Certainly there are key differences between the two industries which are important such as the relatively large size of film files relative to music files, as well as the extent to which a download provides a substitute with the social experience of cinematic consumption. Also, over the time-frame of our study, Australian broadband internet plans and speeds were often restrictive for downloading films but this will change dramatically in the very near future—especially with the roll-out of the National Broadband Network (NBN).

References

Australian Bureau of Statistics (2011). Catalogue 8153.0. Internet Activity Australia, June 2011.

Bhattacharjee, S., Gopal, R., Lertwachara, K., Marsden, J., and Telang, R. (2007). The effect of digital sharing technologies on music markets: A survival analysis of albums on rankings charts. *Management Science*, 53:1359–1374.

Britton, S. (2011). Exclusive—Australian downloaders targeted. *www.mediawave.tv*, March 24. Available at `http://www.mediawave.tv/site/item.cfm?item=F695F6940631EA84670E9237FA46C46E`.

Cragg, J. and Donald, S. (1993). Testing identfiability and specification in instrumental variables models. *Econometric Theory*, 9:222–240.

Danaher, B. and Waldfogel, J. (2012). Reel piracy: The effect of online film piracy on international box office sales. unpublished manuscript, Wellesley Colloege.

De Roos, N. and McKenzie, J. (2014). Cheap Tuesdays and the demand for cinema. *International Journal of Industrial Organization*, 33:93–109.

De Vany, A. and Walls, W. D. (2007). Estimating the effects of movie piracy on box-office revenue. *Review of Industrial Organization*, 30(4):291–301.

Dejean, S. (2009). What can we learn from empirical studies about piracy? *CESifo Economic Studies*, 55:326–352.

Ernesto (2012). Who's pirating Game of Thrones, and why? *www.torrentfreak.com*, May 20. Available at `http://torrentfreak.com/whos-pirating-game-of-thrones-and-why-120520/`.

Hansen, L. P. (1982). Large sample properties of generalised method of moments estimators. *Econometrica*, 50:1029–1054.

Hausman, J., Leonard, G., and Zona, J. (1994). Competitive analysis with differentiated products. *Annals of Economics and Statistics*, 34:159–180.

Hayashi, F. (2000). *Econometrics*. Princeton University Press, Princeton, New Jersey.

Hennig-Thurau, T., Henning, V., and Sattler, H. (2007). Consumer file sharing of motion pictures. *Journal of Marketing*, 71:1–18.

Kleibergen, F. and Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, 133:97–126.

Liebowitz, S. (2006). File-sharing: Creative destruction or just plain destruction? *Journal of Law and Economics*, 49:1–28.

Liebowitz, S. (2007). How reliable is Oberholzer-Gee and Strumpf's paper on file-sharing. unpublished manuscript, University of Texas at Dallas.

Liebowitz, S. (2008). Testing file sharing's impact on music album sales in cities. *Management Science*, 54:853–859.

Liebowitz, S. (2010). The Oberholzer-Gee/Strumpf file-sharing instrument fails the laugh test. unpublished manuscript, University of Texas at Dallas.

McKenzie, J. (2012). The economics of movies: A literature survey. *Journal of Economic Surveys*, 26(1):42–70.

Michel, N. (2006). The impact of digital file sharing on the music industry: An empirical analysis. *The B.E. Journal of Economic Analysis and Policy*, 6(1). Article 18.

Mitra-Kahn, B. H. (2011). Copyright, evidence and lobbynomics: The world after the UK's Hargreaves review. *Review of Economic Research on Copyright Issues*, 8:65–100.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69:304–342.

Oberholzer-Gee, F. and Strumpf, K. (2007). The effect of file-sharing on record sales: An empirical analysis. *Journal of Political Economy*, 115:1–42.

Peitz, M. and Waelbroeck, P. (2004). The effect of internet piracy on music sales: Cross-section evidence. *Review of Economic Research on Copyright Issues*, 1:71–79.

Peitz, M. and Waelbroeck, P. (2006). Piracy of digital products: A critical review of the theoretical literature. *Information Economics and Policy*, 18:449–476.

Rob, R. and Waldfogel, J. (2006). Piracy on the high C's: Music downloading , sales displacement, and social welfare in a sample of college students. *Journal of Law and Economics*, 49:29–62.

Rob, R. and Waldfogel, J. (2007). Piracy on the silver screen. *Journal of Industrial Economics*, 55:379–396.

Stevans, L. and Sessions, D. (2005). An empirical investigation into the effect of music downloading on the consumer expenditure of recorded music: A time series approach. *Journal of Consumer Policy*, 28:311–324.

TERA Consulting (2010). Building a digital economy: the importance of saving jobs in the EU's creative industries. Available at `http://www.droit-technologie.org/upload/dossier/doc/219-1.pdf`.

Waldfogel, J. (2010). Music file sharing and sales displacement in the iTunes era. *Information Economics and Policy*, 22:306–314.

Waldfogel, J. (2012). Music piracy and its effects on demand, supply and welfare. In Lerner, J. and Sten, S., editors, *Innovation Policy and the Economy Volume 12*. University of Chicago Press, NBER.

Walls, W. D. (2008). Cross-country analysis of movie piracy. *Applied Economics*, 40:625–632.

Zentner, A. (2005). File sharing and international sales of copyrighted music: An empirical analysis with a cross section of countries. *The B.E. Journal of Economic Analysis and Policy*, 5(1). Article 21.

Zentner, A. (2006). Measuring the effect of music download on music purchases. *Journal of Law and Economics*, 49:63–90.

Zuel, B. (2012). Australians world's worst for illegal music downloads. *Sydney Morning Herald*, September 19.

Table 1: Film Summary Statistics

Variable	Obs	Mean	Median	Std. Dev.	Min	Max
Total revenue (AUD)	166	8,869,875	6,121,475	9,310,205	559	50,800,000
Total downloads	166	112,971	99,622	63,364	2,816	435,176
Filesize (MB)	166	1,200	1,320	408	637	3,470
Budget (USm)	166	70.3	50.0	56.4	1.5	260.0
Aus opening weekend screens	166	259	242	135	2	758
US opening weekend theatres	166	3,125	3,045	511	2,012	4,468

Table 2: Estimation Sample Summary Statistics — By State

		Revenue					Downloads				
	Obs	Mean	Median	Std. Dev.	Min	Max	Mean	Median	Std. Dev.	Min	Max
State											
ACT	5,938	5,039	2,058	8,667	1	172,402	9.8	7.0	9.6	1	113
NSW	10,077	46,785	6,430	109,982	5	2,486,135	165.1	116.0	167.7	1	2,011
NT	654	3,235	1,571	5,010	16	57,125	1.1	1.0	0.4	1	5
QLD	9,258	31,436	6,774	66,907	5	1,195,300	117.9	80.0	121.4	1	1,271
SA	9,122	9,531	2,096	20,574	8	515,233	53.3	38.0	52.9	1	583
TAS	5,237	3,788	1,610	6,463	8	90,582	13.3	10.0	12.9	1	133
VIC	9,437	41,048	7,658	90,642	1	1,980,287	155.6	108.0	161.4	1	2,132
WA	6,940	19,634	5,530	37,772	5	608,630	78.5	59.0	75.4	1	1,160
All states	56,663	25,148	3,593	69,068	1	2,486,135	95.0	52.0	127.6	1	2,132

Notes: All revenue amounts in Australian dollars.

Table 3: Estimation Results—Contemporaneous Downloads Model

ln Revenue$_{ist}$			GMM	
	OLS	OLS	1st stage	2nd stage
ln Dowloads$_{ist}$	0.150***	0.074***		0.031***
	(0.004)	(0.005)		(0.006)
ln Downloads$_{is't}$			0.959***	
			(0.019)	
File size$_i$			-0.588	
			(0.462)	
Trend$_t$			0.003	
			(0.002)	
ln Theatres$_{ist}$	1.416***	1.103***	0.031***	1.114***
	(0.005)	(0.006)	(0.002)	(0.006)
Week-of-run$_{ist}$	-0.083***	-0.222***	-0.008	-0.216***
	(0.002)	(0.003)	(0.015)	(0.003)
Day-of-week dummies	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes
Film FEs	No	Yes	Yes	Yes
Under Identified			7008.3	
(P-Value)			(0.000)	
Weakly Identified			46697.8	
(P-Value)			(0.000)	
Over Identified			241.7	
(P-Value)			(0.000)	
N	56663	56663	56663	56663
R^2	0.846	0.891	0.942	0.890

Notes: Subscripts i, s and t refer to film, state/territory and date, respectively. Subscript s' refers to all other states/territories $s' \neq s$. Robust standard errors are in parentheses. *, ** and *** denote two tailed significance at 10%, 5% and 1%, respectively.

Table 4: Estimation Results—Dynamic Stock of Downloads Model

ln $Revenue_{ist}$	T=7			T=14		
		GMM			GMM	
	OLS	1st stage	2nd stage	OLS	1st stage	2nd stage
ln $Dowloads_{isT}$	-0.002		-0.062***	-0.107***		-0.166***
	(0.006)		(0.006)	(0.006)		(0.006)
ln $Downloads_{is'T}$		0.955***			0.945***	
		(0.003)			(0.003)	
File $size_i$		-0.438			-0.539*	
		(0.291)			(0.279)	
$Trend_t$		0.003**			0.003***	
		(0.001)			(0.001)	
ln $Theatres_{ist}$	1.108***	0.033***	1.121***	1.121***	0.039***	1.136***
	(0.006)	(0.002)	(0.006)	(0.006)	(0.001)	(0.006)
Week-of-run_{ist}	-0.220***	-0.004	-0.213***	-0.211***	-0.004	-0.202***
	(0.003)	(0.009)	(0.003)	(0.003)	(0.009)	(0.003)
N	56663	56663	56663	56663	56663	56663
R^2	0.890	0.974	0.889	0.891	0.979	0.890
	T=21			T=28		
		GMM			GMM	
	OLS	1st stage	2nd stage	OLS	1st stage	2nd stage
ln $Dowloads_{isT}$	-0.195***		-0.250***	-0.257***		-0.308***
	(0.006)		(0.006)	(0.006)		(0.006)
ln $Downloads_{is'T}$		0.941***			0.940***	
		(0.003)			(0.003)	
File $size_i$		-0.716***			-0.743***	
		(0.281)			(0.275)	
$Trend_t$		0.004***			0.004***	
		(0.001)			(0.001)	
ln $Theatres_{ist}$	1.134***	0.042***	1.150***	1.141***	0.041***	1.155***
	(0.006)	(0.001)	(0.006)	(0.006)	(0.001)	(0.006)
Week-of-run_{ist}	-0.198***	-0.007	-0.188***	-0.186***	-0.006***	-0.175***
	(0.003)	(0.009)	(0.003)	(0.003)	(0.009)	(0.003)
N	56663	56663	56663	56663	56663	56663
R^2	0.893	0.982	0.892	0.893	0.984	0.894

Notes: All models include day-of-week dummies, state fixed-effects, and film fixed-effects. Subscripts i, s and t refer to film, state/territory and date, respectively. Subscript s' refers to all other states/territories $s' \neq s$ and subscript T refers to sum over window $[t, t-T]$ (see text for full details). Robust standard errors are in parentheses. *, ** and *** denote two tailed significance at 10%, 5% and 1%, respectively.

Table 5: Estimation Results—Opening Week Model

ln Revenue$_{ist}$	GMM 2nd stage				
	Contemp.	T=7	T=14	T=21	T=28
ln Downloads$_{is'T}$	-0.028***	-0.062***	-0.050***	-0.048***	-0.051***
	(0.008)	(0.013)	(0.014)	(0.014)	(0.014)
ln Theatres$_{ist}$	1.042***	1.052***	1.055***	1.056***	1.058***
	(0.046)	(0.046)	(0.047)	(0.047)	(0.047)
N	7429	7429	7429	7429	7429
R^2	0.950	0.950	0.950	0.950	0.950

Notes: All models include day-of-week dummies, state fixed-effects, and film fixed-effects. Subscripts i, s and t refer to film, state/territory and date, respectively. Subscript s' refers to all other states/territories $s' \neq s$ and subscript T refers to sum over window $[t, t-T]$ (see text for full details). Robust standard errors are in parentheses. *, ** and *** denote two tailed significance at 10%, 5% and 1%, respectively.

Table 6: Estimation Results—Weekly Model

$\ln$ Revenue$_{ist^W}$	GMM 2nd stage				
	Contemp.	T^W=1	T^W=2	T^W=3	$T^W = 4$
$\ln$ Downloads$_{is'T^W}$	0.284***	0.037**	-0.076***	-0.126***	-0.145***
	(0.021)	(0.018)	(0.016)	(0.015)	(0.015)
$\ln$ Theatres$_{ist^{W(max)}}$	0.986***	1.019***	1.030***	1.031***	1.027***
	(0.014)	(0.015)	(0.015)	(0.015)	(0.015)
Week-of-run$_{ist}$	-0.266***	-0.272***	-0.269***	-0.265***	-0.262***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
N	9228	9228	9228	9228	9228
R^2	0.909	0.898	0.897	0.898	0.898

Notes: All models include day-of-week dummies, state fixed-effects, and film fixed-effects. Subscripts i, s and t^W refer to film, state/territory and date (expressed as week). Subscript s' refers to all other states/territories $s' \neq s$ and subscript T^W refers to sum over window $[t^W, t^W - T^W]$ (see text for full details). Robust standard errors are in parentheses. *, ** and *** denote two tailed significance at 10%, 5% and 1%, respectively.

Table 7: Estimation Results—Forward Dynamic Stock Model

ln Revenue$_{ist}$	GMM 2nd stage				
	Contemp.	T^F=7	T^F=14	T^F=21	T^F=28
ln Downloads$_{is'T^F}$	0.031***	0.096**	0.140***	0.168***	0.181***
	(0.006)	(0.008)	(0.009)	(0.009)	(0.009)
ln Theatres$_{ist}$	1.114***	1.113***	1.115***	1.115***	1.116***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Week-of-run$_{ist}$	-0.216***	-0.215***	-0.214***	-0.214***	-0.213***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
N	56663	56663	56663	56663	56663
R^2	0.890	0.891	0.891	0.892	0.892

Notes: All models include day-of-week dummies, state fixed-effects, and film fixed-effects. Subscripts i, s and t refer to film, state/territory and date (expressed as week). Subscript s' refers to all other states/territories $s' \neq s$ and subscript T^F refers to sum over window $[t, t + T^F]$ (see text for full details). Robust standard errors are in parentheses. *, ** and *** denote two tailed significance at 10%, 5% and 1%, respectively.

Table 8: Sales Displacement Effects

Daily Model	T=7	T=14	T=21	T=28
Estimated Elastisity	-0.062	-0.166	-0.25	-0.308
Adjusted Median Downloads[a]	680	1345	1962	2485
Implied ∂Revenue/∂Downloads[b]	-0.328	-0.443	-0.458	-0.445
Substitution Rate[c]	-2.545	-3.444	-3.558	-3.460
Opening Week Model	T=7	T=14	T=21	T=28
Estimated Elastisity	-0.062	-0.05	-0.048	-0.051
Adjusted Median Downloads[a]	371	565	665	725
Implied ∂Revenue/∂Downloads[b]	-5.190	-2.745	-2.239	-2.183
Substitution Rate[c]	-40.323	-21.330	-17.400	-16.958
Weekly Model	T^W=1	T^W=2	T^W=3	T^W=4
Estimated Elastisity	-	-0.076	-0.126	-0.145
Adjusted Median Downloads[a]	-	1562	1933	2232
Implied ∂Revenue/∂Downloads[b]	-	-0.811	-1.093	-1.093
Substitution Rate[c]	-	-6.300	-8.496	-8.492

Notes: [a] state/territory level median downloads adjusted for Peer Media Technology's estimate of 55% market coverage. [b] calculated as estimated elasticity*(Median Revenue/Median Downloads), where (state/territory level) Median Revenue equals A$3,593, A$31,046, and A$16,771 for the Daily, Opening Week, and Weekly models, respectively. [c] substitution rate divides 'Implied ∂Revenue/∂Downloads' by average ticket price (A$12.87) and multiplies by 100. It provides a measure of the number of admissions displaced by one hundred downloads (evaluated for the median film in terms of revenue and downloads).

Figure 1: Timing of Downloads

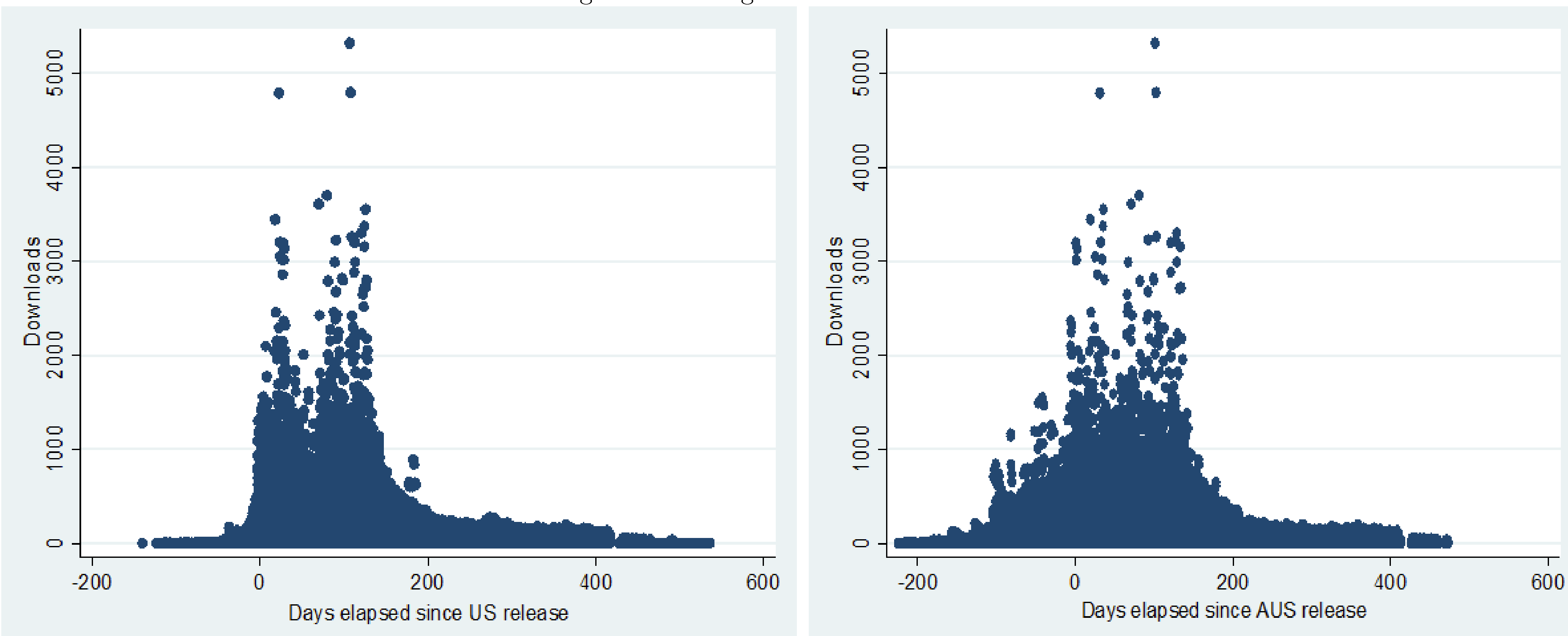

Figure 2: Total Revenues vs. Total Downloads (N=166)