Patent protection, creative destruction, and generic entry in pharmaceuticals: Evidence from patent and pricing data

Alan C. Marco[1]

October 28, 2005

[1]Department of Economics, Vassar College, Poughkeepsie, NY 12604, marco@vassar.edu. This research was supported by a grant from the Ford Foundation. I am also grateful to Merck for data assistance.

Vassar College Economics Working Paper # 83

Abstract

This paper merges patent citation data with data on pharmaceutical patent expirations, generic entry, and pricing to explore the effects of observable patent characteristics on off-patent and on-patnet pharmaceutical pricing. Using a sample of drug patents facing generic entry in the 1990s, I find that the price of branded drugs increased on average in the face of generic entry. Importantly, I find that the number of patent citations that a drug receives from other firms is correlated with a decrease in markup and a decrease in the duration of the markup. Conversely, self-citations are correlated with higher prices and slower decay in prices. The results indicate that patent citations may signal the degree of inter-molecule substitution. And, importantly, self-citations may indicate a degree of cumulative patenting that enables a firm to effectively extend or strengthen the original patent protection. This research takes a step forward in understanding the distinction between "positive" citations and "negative" citations related to creative destruction.

JEL: K11, L11, L65.

Keywords: Patents, Patent Citations, Pharmaceuticals, Pricing, Entry

1 Introduction

In recent years the use of patent citations has proliferated in the economic literature. Patent citations are widely used in examining patent value, firm value, innovative performance, and strategic behavior. Part of the interest in citations is due to the recent availability of patent citation data from the National Bureau of Economic Research (NBER), detailed in Hall, Jaffe and Trajtenberg (2001). That project was undertaken because of a recognition that simple patent counts are noisy measures of innovative output (Trajtenberg 1990).

One of the common interpretations of patent citations is that they are an indicator of patent value or patent quality. Indeed, several papers have shown that citation-weighted patent counts are a better measure of innovative output than straight patent counts (Hagedoorn and Cloodt 2003, Schankerman 1998, Lanjouw and Schankerman 1999).

However, another interpretation of patent citations is that they indicate knowledge flows and spillovers (see Jaffe, Trajtenberg and Henderson (1993) and recent contributions by Moretti (2004) and Hussler (2004) among others). The knowledge flow interpretation indicates that downstream patents may build on work by upstream patents so that innovation is cumulative (Scotchmer 1991). As such, citations may not always measure the current quality or value of the cited patent, but rather the rise of a supplanting technology, as suggested by Schumpeter's creative destruction.

This paper merges data on patent citations with data on pharmaceutical entry and pricing in order to distinguish the value-enhancing and substitution effects signaled by citations. The paper investigates the usefulness of patent statistics as a way to measure inter-molecule substitution and intra-molecule protection by drug pioneers. The results have application outside of pharmaceuticals by demonstrating that the interpretation patent citations must be handled with care. Additionally, this paper adds to the body of work showing strategic patenting behavior on the part of innovators (Podolny, Stuart and Hannan 1996, Grindley and Teece 1997, Hall and Ziedonis 2001).

Previous pharmaceutical studies have paid great attention to the protection afforded by patents. Pharmaceutical products are undoubtedly fertile markets in which to investigate the effect of patent protection by comparing on-patent and off-patent pricing. Unlike markets like electronics, patented pharmaceutical products are fairly well-defined, and generic entry is relatively easy to measure. Despite being the very basis of market power in phar-

maceuticals, little attention has been paid to the observable characteristics of the patents themselves. Part of the reason is certainly the data requirements: to examine the effects of expiring patents, one must include in the sample patents granted 20 years prior. Rich patent data are available only since the mid-1970s. Thus, until recently, it was not feasible to examine detailed patent statistics in combination with post-patent pricing and entry.

Several papers have examined the effect of generic entry on pioneer drug pricing after patent protection expires. The empirical results indicate that pioneer prices may either rise or fall in the face of generic competition. For instance, both Grabowski and Vernon (1992) and Frank and Salkever (1997) find that branded prices increase after generic entry. In contrast, Caves, Whinston and Hurwitz (1991) and Wiggins and Maness (1994) find a negative impact of generic competition on branded price. The accepted explanation for branded price increases lies in market segmentation: when faced with generic competition, pioneers may either compete against the generics for price sensitive consumers, or they may forego the price sensitive segment of the market in favor of selling to brand-loyalists. Thus, depending on the strategy that the pioneer firm employs, we can expect to see prices either decrease or increase. In reality, many pioneer firms begin manufacturing and selling their own generic versions of the drug in order to price discriminate based on brand loyalty. In any case, while pioneer prices may rise or fall when faced with generic entry, a price premium is generally enjoyed by the pioneer. observe a negative price response. The accepted theoretical explanation for price increases is that branded drug manufacturers may practice price discrimination after patent expiration: rather than focus on price-sensitive consumers, they restrict their sales to brand loyal consumers. Indeed, many manufacturers produce their own generic versions of their drugs after generic entry—competing on price with generics, and reserving the brand name product for brand loyalists. In any case, whether prices increase or decrease in response to generic entry the branded drug enjoys a price premium over generics, and pioneer profits decrease.

I examine on- and off-patent pricing using a sample of drugs facing generic entry in the 1990s. The paper provides two primary contributions to the existing literature on pharmaceutical entry. First, I exploit the observable characteristics of the patents that were the basis of monopoly power and brand recognition for the drugs in my sample. Second, I explore the dynamics drug pricing using a hazard specification.

I find that the price of branded drugs in my sample increased on average in the face of

generic entry. Additionally, I find that the number of patent citations that a drug receives from other firms is correlated with a decrease in markup, and a decrease in the duration of the markup. Conversely, the number of patent citations that a manufacturer makes to its own drug patent ("self-citations") is correlated with higher prices and slower decay in prices. The results imply that forward citations by other firms signal the degree of inter-molecule substitution, and self-citations correspond to the degree a pioneer is able to further protect its original substance by "fencing" the patent in with follow-on patents. The follow-on patents generally cover new forms of the substance (e.g., oral versus injectible forms or extended release capsules), but may also include new processes that enable cheaper production of the substance. Thus follow-on patenting softens the blow of the patent expiration on the molecule. This research takes an important step in understanding strategic cumulative patenting, as well as understanding the distinct differences between forward citations and self-citations.

In the following section, I present the econometric specification. Next, I describe the pharmaceutical data and patent data in my sample. Section (4) presents the results of the estimations, and Section (5) concludes and discusses possible extensions.

2 Specification

My specification utilizes a simple product differentiation framework in order to estimate a relationship between branded drug price and generic entry and substitution. From the first order conditions for profit maximization for a given drug at a particular time, we get the Lerner equation

$$L = \frac{s}{\varepsilon}, \tag{1}$$

where $L = \frac{p-c}{p}$ is the Lerner index, p and c are the price and marginal cost, respectively, s is the market share relative to generic production, and ε is the molecule-specific demand elasticity. In the case of pharmaceuticals, the pioneer manufacturer generally has some brand name recognition that differentiates it from generic competitors, even after patent expiration.

I assume an inverse demand function of $p = p(q + \theta g)$ where q is the quantity of the branded drug, g is the quantity of the generic competitor, and θ represents the degree of substitutability between branded and generic drugs. I also allow for conjectural variations,

so that $\frac{dg}{dq} = \lambda$. In this case, the Lerner equation becomes

$$L_{it} = \frac{(1 + \theta_{it}\lambda_{it})}{\varepsilon_{it}}. \tag{2}$$

To develop an econometric specification for drug i at time t, I take logs and add an error term u_{it}:

$$\ln L_{it} = \ln(1 + \theta_{it}\lambda_{it}) + \ln \varepsilon_{it} + u_{it}. \tag{3}$$

The first term on the right hand side of Equation (3) measures the degree of substitutability with generic manufacturers, or intra-molecule competition. And the second term represents factors that influence the firm-specific elasticity of the drug, including the availability of therapeutic substitutes (substitute pharmaceutical compounds—inter-molecule competition). My estimation equation is a linear approximation,

$$\ln L_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_{it} + u_{it}, \tag{4}$$

where X_{it} is a matrix of explanatory variables influencing intra-molecule competition, Z_{it} is a matrix of explanatory variables influencing inter-molecule competition. X_{it} and Z_{it} vary somewhat across specifications in Section (4), but the primary variables are discussed here.

For intra-molecule competition, the most important explanatory variable is the number of generic entrants since patent expiration. As discussed below, I do not have direct observations on the number of generic competitors or market shares. Instead I use the number of generic applications (Amended New Drug Applications—ANDAs) at the Food and Drug Administration. Generic producers are required to submit an ANDA prior to generic manufacture. I treat the cumulative total as a proxy for the degree of potential entry. To the extent that entry is credible, the number of potential entrants may be a better measure of competitiveness than observed entry.

For inter-molecule competition, the most important explanatory variable is the number of forward citations received by the patent, discussed in more detail below. I classify forward citations into two types: self-citations and citations by others. Self-citations are calculated from the patent data by multiplying the cumulative forward citation count by the proportion of self-citations. Other citations are the residual. I hypothesize that the number of citations by others (and the technological closeness of those citations) indicates the number of potential inter-molecule substitutes, which should increase the molecule specific demand elasticity, ε_{it}.

In calculating the Lerner index, it is necesary to approximate marginal cost. In my sample, I approximate marginal cost for each branded drug by using the lowest observed generic price for the drug. The absence of any real cost data is a limitation, but if generics are competitive and if manufacturing processes are similar, then the lowest observed generic price during the sample period should well approximate marginal cost. While it might be expected that the costs of production will change over the life cycle of the drug, it is unlikely that they will change significantly in the neighborhood of the patent expiration; little R&D goes into improving productivity for drugs that are nearing expiration.

3 Data

3.1 Drug Sample

My sample of drugs comprises a subset of the drugs found in the Generic Spectra database from IMS Health. Generic Spectra is a database of over 100 drugs facing patent expiration and generic entry between 1992 and 2002. The Generic Spectra drugs are listed in Table (1), along with the availability of entry, price, and patent data. I obtained branded drug names, molecule names, and patent expiration dates from these data. Patent protection on pharmaceuticals can vary, and can be supplemented by "exclusivity" granted by special legislation, e.g., the Drug Price Competition and Patent Term Restoration Act (DPCPTR). The period of exclusivity may or may not outlast patent protection. The Generic Spectra expiration dates are intended to indicate the earliest date at which generics may enter the market.

The patents for the Generic Spectra drugs were obtained from the IMS Patent Focus database. I was able to identify a patent number for 68 pioneer drugs. It should be noted that a drug may be covered by multiple patents; so, for the purposes of obtaining patent statistics, I chose the earliest of issued patents. The earliest patent is usually the patent on the molecule that defines the drug. Later patents are frequently sought on such things as drug delivery or form. The usual practice is for subsequent patents on a drug to cite back to the primary molecule patent. An expired drug may allow for generic competition in its original form; however, the pioneer firm may patent an extended release version which would not immediately be subject to generic competition.

Generic Spectra includes only drugs that experienced some generic entry. These are

usually drugs with larger markets. Throughout, my results are conditional on some amount of entry. So, my results on entry relate to the intensity or frequency of entry, not the likelihood of entry. However, to the extent that the economically important drugs are those that face generic entry, the restriction is not significant.

Table (2) describes the variables used in the study, including some summary statistics.

3.2 Entry

Data on entry were obtained from the FDA's Orange Book. Each generic entrant for a given chemical compound must file an ANDA with the FDA. Filing an ANDA is a necessary but not sufficient condition for entry. In reality the number of entrants represents potential entry, and will differ from actual entry because I do not observe whether the firm actually produces or whether it produces and then exits. But, as mentioned above, to the extent that entry is easy for a generic manufacturer, it may be that potential entry is a better measure of competitive pressure than observed entry.

3.3 Prices

Prices for the drugs in the sample were obtained from the ReadyPrice database from Thomson-Micromedex. The prices listed in this product are Average Wholesale Prices (AWP), which do not reflect any discounts off the list price. Discounts are common in wholesale pharmaceutical markets, so that AWP is referred to as "ain't what's paid." However, as other researchers have noted, as long as AWP is correlated with the actual discounted price, the qualitative results in my estimation should not be influenced by the use of this proxy (Lu and Comanor 1998). AWP prices are used because they are much more readily available; revenue and quantity data are expensive to obtain from private data vendors.

Unfortunately, the pricing data are far from complete. The pricing history is not rich for most drugs, and for many drugs I could not find pricing data for either the pioneer, generics, or both. For simplicity, I restricted pricing data to drugs sold in tablet and capsule form. This restriction eliminated a few drugs, but enabled me to more easily calculate comparable prices based on per unit weight. Because I do not have data on the quantity sold in different forms, some simplification was necessary.

On a given day, A manufacturer may set different prices based on form, dosage, and package size. Where multiple prices exist on a given day, I calculate a simple average unit

price[1] and a minimum unit price observed on that day.

To calculate a Lerner index, I approximate marginal cost using the minimum price observed by a generic entrant over the sample period. Thus, I implicitly assume that marginal cost remains the same over the time period. Again, this assumption is reasonable if we restrict ourselves to the latter part of a patent's life. The earliest price data in the sample come from seven years prior to expiration, and the latest prices are 10 years after expiration.

3.4 Patent statistics

Patent statistics were obtained from the NBER Patent Citations Data Files (described in Hall et al. (2001)).

Of primary importance for the current paper are the citation variables. So-called "forward citations" are citations received by the patent from subsequent patents. Forward citations are frequently associated with higher patent value. The rationale is that if a patent is frequently cited, then it may be the basis for cumulative innovation, and may be technologically important. However, higher forward citations may also have a negative impact on value if citations reflect replacement by new technologies.

In the pharmaceutical context, it is likely that both effects occur. More citations will occur for blockbuster drugs, and they will also occur in crowded therapeutic classes. However, with regard to pricing in particular, I posit that more citations by other firms in nearby technological classes reflect the degree of substitutes that exist for this drug. If we define blockbuster drugs by the size of the market, then we can expect more citations to reflect larger revenues. But, this does not imply that the Lerner index will necessarily be higher.

To determine the effect on price, it is useful to distinguish between forward citations by others, and self-citations. The NBER dataset defines two self-citation variables, reflecting the proportion of forward citations that are made to the patent by the same patent holder. Since information about the patent holder of citing patents is subject to error, the two self-citation variables represent the upper and lower bounds (Hall et al. 2001). Missing values for the self-citation variables will obtain whenever there are no forward citations. In the estimation, I use the lower bound estimate, and replace missing lower bound values with

[1] Obviously, if one had revenues, one would prefer to calculate a weighted average rather than a simple average.

zero in the event of zero citations.

I use the proportion of self-citations to impute the quantity of self-citations. Self-citations represent investment by a firm in a patented line of drugs or related drugs. These citations indicate either an active effort to protect patent coverage (and thus make substitution harder), or they indicate a response by the firm to strong patent protection. That is, the causality is not clear *a priori*, but in either case, self-citations are correlated with more freedom from inter-molecule competition. Similarly, the quantity of other-citations indicates that other firms are citing the patent, and potentially providing substitute drugs or treatments. Both self-citations and other-citations are tracked dynamically. That is, I track when each citation is made, and keep a cumulative total based on date.

By way of example, Table (3) lists all the patent citations made to patent number 4,267,320, which protected the branded drug Ceftin (cefuroxime axetil) manufactured by Glaxo. One can observe that the first several patent citations are self-citations covering new forms of the drug. Glaxo continually introduced new forms (e.g., a solid oral form of the drug rather than the original injectible) and finally a process patent. Sumitomo's citation represents a new chemical that may operate as a therapeutic equivalent to cefuroxime axetil. Notably, there is a gap in citations that resumes when the patent expires. After expiration, one can see a flurry of process patents related to manufacturing for the generic producer Ranbaxy.

Backwards citations, or citations made by the patent to prior patents may also indicate more substitutes in the therapeutic class. In this case, the citations might not be negative in that the new drug may usurp the cited drug. So, this substitution effect is likely to be smaller. Self-citation data are not available easily available for backwards citations.

Hall et al. (2001) describe the creation of two indices in the NBER data using patent citations: *generality* and *originality* (Jaffe et al. 1993). The NBER indices are useful for several reasons. First, they are convenient and available. Additionally—and because of this—they are being used more frequently in empirical research, so they serve as a useful benchmark.

The generality index for patent i is defined as

$$1 - \sum_{j}^{n_i} s_{ij}^2$$

where s_{ij} refers to the proportion of citations to patent i from patents in technology class j.

Thus, the higher the index, the more spread out are the patents that cite it, technologically speaking (Hall et al. 2001). Low generality should be bad news for a branded drug, as it will indicate citations from a more "focused" technology class, making it more likely that it is receiving citations from closer substitutes.

Originality is similarly defined, except that s_{ij} refers to backwards citations (citations *made by* the patent in question) rather than forwards citations (citations *received by* the patent in question). Higher originality indicates more diffuse backward citations technology classes, implying (perhaps) that a wider array of technologies were utilized in the innovation.

Both generality and originality are undefined if the number of citations is zero. In my sample, undefined values are replaced by zero. In the case of generality this redefinition is justifiable in that an uncited patent is not applicable to *any* patented technologies (yet), and therefore receives a low score for generality.[2] In the case of originality, one could imagine that highly original patents might cite *no* previous patents. However, from an empirical standpoint, Hall et al. (2001) observe that higher numbers of citations tend to be associated with higher originality and generality indices; thus, assigning zero to undefined values seems the logical choice.

4 Estimation

In the following subsections, I estimate different versions of Equation (4). In Section (4.1) I estimate ordinary least squares estimates of the price equation, and test for endogeneity. In Section (4.2) I attempt to control for endogeneity using instrumental variables and simultaneous equations methods. Lastly, in Section (4.3), I estimate a hazard rate specification of the likelihood of branded price decreases.[3]

4.1 Price equations

Table (4) shows the results of estimating Equation (4) using ordinary least squares. The dependent variable is the log of the Lerner index. In columns (1) to (4), the branded price used to calculate the Lerner index is the minimum branded price on any given date (L); in the last specification the average price is used ($Lavg$).

[2] Alternatively, the patent may just be very young.

[3] For the sake of parsimony, I will focus my remarks on the primary variables of interest: price, entry, and forward citations.

The first noticeable result is that entry is positively correlated with markups. There are two potential explanations for this result. First, if firms are attempting to segment the market by brand loyalty—as discussed in Section (1)—then we would expect to see price respond positively to entry. However, another possible explanation is endogeneity. If high prices induce entry, then the positive relationship may result from the bias of a simultaneously determined system. I will examine endogeneity below.

Throughout all the estimated OLS equations, the log of the number of forward citations (plus one) by "others" enters significantly. In columns (1) and (2) of Table (4), $\ln(subfor)$ is positive, and in columns (3) to (5)—when dummies for the therapeutic categories have been added—the sign becomes negative and significant at the 1% level. The negative sign is evidence that on average forward citations by others represent potential substitute products, since greater substitution will lead to lower prices. For this reason I refer to citations by others as $subfor$ or "substitute" forward citations.

The sign on the log of the number of self-citations (plus one)—$\ln(selfor)$—is positive throughout, and significant at the 1% level when the category dummies are included. The positive coefficient is consistent with the interpretation of self-citations as representing barriers to (substitute) entry, as discussed in Section (3).

Lastly, *generality* has a significant positive impact on the markup when the categorical dummies are included. A more general patent indicates that forward citations come from a wide array of technology classes. If citations tend to be focused from only one class, it is more likely that these citations are from the same technological class as the drug itself, representing greater substitutability. Thus, higher generality (less focus) should indicate less substitution, and a higher price. In fact, the coefficients on $\ln(subfor)$ and *generality* must be interpreted together: $\ln(subfor)$ represents the amount of (potentially) substitute citations, and *generality* represents the intensity (or lack of intensity) in a particular technological class.

Because of possible endogeneity between prices and entry, the OLS coefficients cannot be relied upon. A Durbin–Wu–Hausman test (described in Davidson and MacKinnon (1993)) reveals the presence of endogeneity in the OLS estimates. I regress $\ln(entrants)$ on the log of time since patent expiration ($\ln(t)$), patent characteristics, and the categorical dummies. The residual from that regression is included in a price equation (column(3) of Table(4)) to test for endogeneity. The coefficient on the residual is significant at the 5% level, indicating

endogeneity of entry. I employ two approaches to control for the problem: instrumental variables and simultaneous equations.

4.2 Endogeneity

Table (5) shows the results of the instrumental variables approach, when I instrument with $\ln(t)$. I treat the time since expiration as the identification variable because it is more likely that some entry (generic and substitute) occurs over time, and that the increased entry impacts price. However, it is unlikely that the passage of time impacts price in a distinct way.[4]

The four columns of Table (5) represent specifications using both $\ln(L)$ and $\ln(Lavg)$ as dependent variables. Columns (2) and (4) include category dummies.

Entry fails to be significant in any estimation, but the coefficients on $\ln(subfor)$, $\ln(selfor)$, and *generality* continue to be highly significant when category dummies are included. The signs remain the same as in the OLS specification.

Table (6) shows the results of two simultaneous equations models. Again, I use both $\ln(L)$ and $\ln(Lavg)$ as measures of the markup; these are dependent variables in the price equations and independent variables in the entry equations. The price equations (columns (1) and (3)) are identified by $\ln(subfor)$ and $\ln(selfor)$. The entry equations are identified by $\ln(t)$ and $offpat$ (an indicator variable equal to one when the drug is off-patent).

In the price equations, the signs and significance remain the same as in the OLS specifications. That is, entry again enters positively, and the signs on $\ln(subfor)$, $\ln(selfor)$, and *generality* continue to be negative, positive, and positive, respectively. In the entry equation, the markup has a positive impact on entry, which is what we should expect, theoretically. Unreported estimates using the generic markup as an additional explanatory variable in the entry equation do not alter the coefficients on the citation variables. The results of the simultaneous equations models support other studies that find a positive price response to entry.

In summary, I find my hypotheses confirmed. In particular:

- Self-citations tend to increase the branded markup.

[4] Clearly it can be argued that drug-specific demand elasticity will be larger over time. However, this substitution is likely to take place over the 10 to 15 year that the drug is on-patent. The elasticity in the post-patent timeframe is likely to be fairly constant.

- Forward citations by others tend to decrease the branded markup.
- Entry is met by an increase, on average, of the branded markup.
- Generality (less intense technological focus of forward citations) tends to increase the markup.

Besides the size of the markup, it is interesting to investigate the dynamics of the branded response to generics. To examine the duration of the markup following entry, I utilize a duration model.

4.3 Duration

I begin the duration—or hazard rate—analysis by specifying a reduced form model for the probability of a decrease in the branded price. The hazard function, $\lambda(t)$, gives the probability that the pioneer firm will decrease its price given that it has not decreased its price in the previous t years. The hazard function is defined as $\lambda(t) = \frac{f(t)}{1-F(t)}$, where $f(t)$ and $F(t)$ are the usual density and cumulative probability functions.

The exponential specification assumes a constant hazard: $\lambda(t) = \lambda$, so that the hazard function does not vary with the duration of the spell. That is, there is no duration dependence; the length of time a firm has gone without lowering price does not, *ceteris paribus*, affect the likelihood of a price decrease in the next interval of time. The hazard rate is constant in t if the corresponding distribution is exponential.

The Weibull distribution leads to a hazard function of the form $\lambda(t) = \lambda p(\lambda t)^{p-1}$. This hazard function includes the exponential as a special case where $p = 1$, therefore it is useful to include it as a comparison. For values of $p < 1$, the hazard function will be decreasing in time (it will exhibit negative duration dependence). For $p > 1$, the hazard function will exhibit positive duration dependence. For both the exponential and Weibull models, the parameter λ is modeled as

$$\lambda = e^{X\beta+\varepsilon} \tag{5}$$

where X is a matrix of drug and patent characteristics given in Table (2). Some of the covariates vary over time: notably the current number of entrants, and the number of substitute citations and self-citations.

Estimation involves maximum likelihood estimation where the censored observations are incorporated into the log-likelihood function much like the Tobit model (Greene 1993):

$$\sum_{uncensored} \lambda\left(t|\theta\right) + \sum_{all} \ln\left(1 - F\left(t|\theta\right)\right) \tag{6}$$

Table (7) show the results of the hazard estimation for six specifications. The coefficients are presented in their exponentiated forms (e^{β}) so that a value above one indicates a positive marginal effect, and a value below one indicates a negative marginal effect. Columns (1) to (3) assume an exponential distribution, and columns (4) to (6) assume a Weibull distribution. For each specification I use three samples: all available time periods, the on-patent subsample, and the off-patent subsample. For the on-patent subsamples, generic entry cannot be used as an explanatory variable.

All specifications show the same impact of the relevant explanatory variables. The impact of forward citation variables are consistent with those found in the price equations in Sections (4.1) and (4.2). The impact of entry is more ambiguous.

The primary findings are:

- More entry and more substitute citations increase the likelihood that prices will fall.
- More self-citations and higher generality decrease the likelihood that prices will fall.
- The impact of substitute citations and generality is much stronger in the on-patent sample than in the off-patent sample.

Regarding entry, it may seem odd that it would lead to higher price markups and faster price decreases. In an unreported estimation, I perform the same analysis on price *increases* and find that more entry also leads to an increase in the likelihood that prices will increase. The explanation is that some drugs will decrease price in the face of generic competition, and some drugs will increase price;[5] regardless of which strategy the firm follows, it will follow it faster if there is more entry.

The information regarding forward citations fills out the picture of the substitution story. More substitute citations and lower generality will lead to lower prices and speedier price decreases, while more self-citations will lead to higher prices and a lower likelihood of price decreases. The interpretation of substitute citations and self-citations as relating to

[5] On average prices increase—as found in Sections (4.1) and (4.2). But, there is heterogeneity in response.

inter-molecule substitution more than *intra*-molecule substitution is bolstered by the fact that the coefficients are much more extreme in the on-patent sample (where there is no generic competition) than in the off-patent sample.

A final interesting result is that the scaling parameter in the Weibull specifications indicates that there is positive duration dependence. That is, the longer an off-patent drug goes without lowering price, the *less* likely that it is to lower price in the future.

5 Conclusion

The results presented above are important for the pharmaceutical literature because they suggest that citations made to the pioneer drug patent by other firms may be related to the degree that other products (molecules) compete with the drug. The effects of the *generality* index also seems to indicate that more focused forward citations are correlated with the existence of closer substitutes. Similarly, self-citations may indicate barriers to entry (or at least strong patent protection), and lead to higher prices. I also find that pioneer drug prices tend to increase when faced with generic entry, supporting the market segmentation theories of other researchers. Using hazard rate analysis, I show that price decreases by pioneer firms exhibit negative duration dependence.

Additionally, the results are important for the patent literature. Researchers have long been aware that patent citations are a way to measure patent quality or value. However, they are equally aware that citations may represent a negative impact relative to self-citations in the form of supplanting technology. Schumpeter's creative destruction, in fact, relies on patent citations being a negative indicator, at least on some level. This paper is the first to my knowledge to distinguish empirically between negative citations and positive citations. The pharmaceutical industry offers a unique opportunity to study citation, because it is easier to distinguish positive from negative citations, due to the nature of the technology and the nature of competition.

Lastly, the methodology emphasizes the usefulness of citation data outside of their usual applications. They are widely available and cheap. So, to the extent that they proxy well for different competitive phenomena—like entry and strategic investment—they can be utilized as proxies in many contexts.

One caution should be noted in interpreting the results. First, while the interpretation

of the coefficient estimates is consistent with a substitution story, there is more information available in the citation data to exploit. Detailed coding of citations, the technologies involved (compound claims, composition claims, method-of-use claims, and process claims), and tracking specifically who cites whom (and why) is a fruitful area for ongoing research.

Extensions in better exploiting patent data should prove profitable and this paper takes a step in bridging the gap between the empirical patent literature and the pharmaceutical pricing literature. Patent statistics are a potentially valuable—and largely unexploited—resource for investigating pricing and competition.

References

Caves, Richard E., Michael D. Whinston, and Mark A. Hurwitz, "Patent Expiration, Entry, and Competition in the U.S. Pharmaceutical Industry," *Brookings Papers on Economic Activity (Microeconomics)*, 1991, pp. 1–48.

Davidson, R. and J. G. MacKinnon, *Estimation and Inference in Econometrics*, New York: Oxford University Press, 1993.

Frank, Richard G. and David S. Salkever, "Generic Entry and the Pricing of Pharmaceuticals," *Journal of Economics & Management Strategy*, Spring 1997, *6* (1), 75–90.

Grabowski, Henry G. and John M. Vernon, "Brand Loyalty, Entry, and Price Competition in Pharmaceuticals After the 1984 Drug Act," *Journal of Law and Economics*, October 1992, *35* (2), 331–50.

Greene, William H., *Econometric Analysis, 2nd edition*, New York: MacMillan, 1993.

Grindley, Peter C. and David J. Teece, "Managing intellectual capital: Licensing and cross-licensing in semiconductors and electronics," *California Management Review*, Winter 1997, *39* (2), 1–34.

Hagedoorn, John and Myriam Cloodt, "Measuring Innovative Performance: Is There an Advantage in Using Multiple Indicators?," *Research Policy*, September 2003, *32* (8), 1365–79.

Hall, Bronwyn, Adam B. Jaffe, and Manuel Trajtenberg, "The NBER Patent Citation File: Lessons, Insights and Methodological Tools," *NBER Working Paper Number 8498*, October 2001.

Hall, Bronwyn H. and Rosemarie Ham Ziedonis, "The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979-1995," *RAND Journal of Economics*, Spring 2001, *32* (1), 101–128.

Hussler, Caroline, "Culture and Knowledge Spillovers in Europe: New Perspectives for Innovation and Convergence Policies?," *Economics of Innovation and New Technology*, September 2004, *13* (6), 523–41.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson, "Geographic localization of knowledge spillovers as evidenced by patent citations," *Quarterly Journal of Economics*, 1993, *108* (3), 577–598.

Lanjouw, Jean O. and Mark Schankerman, "The Quality of Ideas: Measuring Innovation with Multiple Indicators," *National Bureau of Economic Research Working Paper 7345*, 1999.

Lu, Z. John and William S. Comanor, "Strategic Pricing of New Pharmaceuticals," *Review of Economics and Statistics*, February 1998, *80* (1), 108–118.

Moretti, Enrico, "Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions," *American Economic Review*, June 2004, *94* (3), 656–90.

Podolny, Joel M., Toby E. Stuart, and Michael T. Hannan, "Networks, knowledge, and niches: Competition in the worldwide semiconductor industry, 1984-1991," *American Journal of Sociology*, November 1996, *102* (3), 659–689.

Schankerman, Mark, "How valuable is patent protection? Estimates by technology field," *Rand Journal of Economics*, Spring 1998, *29* (1), 77–107.

Scotchmer, Suzanne, "Standing on the shoulders of giants: Cumulative research and the patent law," *Journal of Economic Perspectives*, Winter 1991, *5* (1), 29–41.

Trajtenberg, Manuel, "A Penny for Your Quotes: Patent Citations and the Value of Innovations," *RAND Journal of Economics*, Spring 1990, *21* (1), 172–87.

Wiggins, Steven N. and Robert Maness, "Price Competition in Pharmaceutical Markets," 1994. Unpublished Manuscript, Department of Economics, Texas A&M University.

6 Tables

Table 1: List of sample drugs

Category	Pioneer Drug	Molecule	Patent data	Generic prices	Branded prices	Entry data	All
analgesics	dolobid	diflusinal		x	x	x	
	stadol	butorphanol tart	x			x	
	toradol	ketorolac	x	x	x	x	x
	ultram	tramadol		x		x	
anesthetics	amidate	etomidate				x	
	forane	isoflurane	x			x	
	sufenta	sufentanil	x				
antiarthritics	ansaid	flurbiprofen	x	x	x	x	x
	daypro	oxaprozin		x	x	x	
	lodine	etodolac	x	x	x	x	x
	naprosyn	naproxen	x	x	x	x	x
	orudis	ketoprofen	x	x	x	x	x
	relafen	nabumetone	x	x	x	x	x
	voltaren	diclofenac sodium	x			x	
antibiotics/anti-infectives	amikin	amikacin	x			x	
	ceclor	cefaclor	x	x	x	x	x
	ceftin	cefuroxime axetil	x	x	x	x	x
	mefoxin	cefoxitin				x	
	zinacef	cefuroxime	x			x	
antidepressants	anafranil	clomipramine		x		x	
	aventyl	nortriptyline	x	x		x	
	luvox	fluvoxamine	x	x		x	
	prozac	fluoxetine	x	x	x	x	x
	vivactil	protriptyline			x		
	wellbutrin	bupropion	x	x	x	x	x
antifungal	lotrimin	clotrimazole				x	
	nizoral	ketoconazole	x	x	x	x	x
antimalarial	plaquenil	hydroxycholoroquine		x	x	x	
antivirals	flumadine	rimantadine		x	x	x	
	zovirax	acyclovir	x	x	x	x	x
bile therapy	actigall	ursodiol		x	x	x	
cancer	blenoxane	bleomycin				x	
	eulexin	flutamide	x	x	x	x	x
	hydrea	hydroxyurea		x	x	x	
	mutamycin	mitomycin				x	
	nolvadex	tamoxifen		x	x	x	
	taxol	paclitaxel				x	
	vepesid	etoposide				x	
cardiovascular	betapace	sotalol HCL		x	x	x	
	bumex	bumetanide	x	x	x	x	x
	capoten	captopril	x	x	x	x	x
	capozide	captopril	x				
	cardene	nicardipine	x	x		x	
	cardizem	diltiazem				x	
	cardura	doxazosin mesylate	x	x	x	x	x
	corgard	nadolol	x	x	x	x	x
	dobutrex	dobutamine	x			x	
	hytrin	terazosin	x	x	x	x	x
	lopid	gemfibrozil	x	x	x	x	x
	lopressor	metoprolol	x			x	
	lozol	indapamide		x	x	x	
	mevacor	lovastatin	x	x	x	x	x
	mexitil	mexiletine	x	x	x	x	x
	primacor	milrinone	x			x	
	prinivil	lisinopril	x	x	x	x	x
	procardia	nifedipine	x			x	
	questran	cholestyramine				x	
	rythmol	propafenone HCL		x	x	x	
	sectral	acebutolol	x	x	x	x	x
	tambocor	flecainide	x	x	x	x	x
	tenex	guanfacine	x	x	x	x	x
	tenoretic	atenolol/chlorthalidone	x	x		x	
	vasotec	enalapril maleate	x	x	x	x	x
	visken	pindolol		x		x	
	wytensin	guanabenz		x		x	
	zebeta	bisoprolol	x	x	x	x	x
	ziac	bisoprolol	x	x	x	x	x

Category	Pioneer Drug	Molecule	Patent data	Generic prices	Branded prices	Entry data	All
dermatological	temovate	clobetasol propionate	x			x	
diabetes	glucophage	metformin		x		x	
	glucotrol	glipzide	x	x	x	x	x
gastrointestinal	axid	nizatidine	x	x	x	x	x
	carafate	sucralfate		x	x	x	
	pepcid	famotidine	x	x	x	x	x
	prilosec	omeprazole	x			x	
	tagamet	cimetidine	x	x	x	x	x
	zantac	ranitidine	x	x		x	
hemotological	coumadin	warfarin sodium		x	x	x	
	trental	pentoxifylire				x	
inmunological	imuran	azathioprine		x	x	x	
	neoral	cyclosporine	x	x		x	
musculoskeletal	aredia	pamidronic acid	x			x	
	tracrium	atracurium besilate	x			x	
	zanaflex	tizanidine	x	x		x	
neurological	eldepryl	selegiline	x	x	x	x	x
	klonopin	clonazepam				x	
	parlodel	bromocriptine	x	x	x	x	x
	permax	pergolide		x	x	x	
	sinemet	carbidope/levodopa		x		x	
	zarontin	ethosuximide				x	
ob/gyn	estrace	estradiol		x	x	x	
	ogen	estropipate		x	x	x	
opthalmics	betagan	levobunolol	x			x	
	betoptic	betaxolol	x			x	
	neptazane	methazolamide		x		x	
	propine	dipivefrine	x			x	
psychotherapeutics	buspar	buspirone	x	x		x	
	clozaril	clozapine		x	x	x	
	cylert	pemotine				x	
	versed	midazolam HCL	x			x	
	xanax	alprazolam	x			x	
respiratory	atrovent	ipratropium	x			x	
	intal	cromolyn	x			x	
	seldane	terfenadine	x				
sedatives	halcion	triazolam	x			x	
	prosom	estazolam	x			x	
tuberculosis	rifadin	rifampin		x	x	x	
		Counts	68	67	53	102	35

Table 2: Variables used in estimation

Prices and entry			Obs	Mean	Std. Dev.
	pbmin	The minimum brand price on a given date.	504	2.09	1.40
	pbavg	The average brand price on a given date.	504	2.28	1.57
	pbdown	Indicator variable = 1 if there is a branded price decrease.			
	pgmin	The minimum generic price on a given date.	1235	1.33	0.89
	pgavg	The average generic price on a given date.	1235	1.52	1.00
	cost	The minimum of any generic price ever facing the pioneer drug.	3006	0.66	0.70
	L	Branded Lerner index: (pbmin - cost)/cmin.	503	0.60	0.29
	Lavg	Branded Lerner index: (pbavg - cost)/cmin.	503	0.63	0.26
	entrants	Cumulative number of generic applicants.	3829	4.37	6.02
Patents					
	subfor	Cumulative number of forward citations made by other firms.	3304	34.22	38.51
	selfor	Cumulative number of forward citations made by own firm.	3304	3.71	8.77
	general	NBER "Generality" index. Undefined values set to 0	3304	0.53	0.16
	original	NBER "Originality" index. Undefined values set to 0	3304	0.19	0.29
	cmade	Number of citations made by the patent.	1982	2.95	2.18
	offpat	Indicator variable = 1 when the patent has expired.			
Other					
	cat	dummy variables for pharmaceutical category			
	t	time since patent grant			

Table 3: Citations to patent 4,267,320 (Ceftin/Glaxo)

Citing Patent	Patent Assignee	Form	Process	Chemical
4385054	Glaxo	X		
4446317	Glaxo	X		
4562181	Glaxo	X		
4602012	Glaxo	X		
4705784	Sumitomo			X
4820833	Glaxo	X		
4897270	Glaxo	X		
4994567	Glaxo		X	
6060599	Ranbaxy		X	
6323193	Ranbaxy	X		
6346530	Ranbaxy		X	
6384213	Ranbaxy		X	
6485744	Individual	X		
6534494	Ranbaxy		X	
6833452	Ranbaxy		X	

Table 4: Ordinary least squares estimation

	(1) ln(L)		(2) ln(L)		(3) ln(L)		(4) ln(L)		(5) ln(Lavg)	
ln(entrants)	0.238	***	0.247	***	0.181	***	0.195	***	0.184	***
	(0.030)		(0.031)		(0.026)		(0.027)		(0.024)	
ln(subfor)	0.050	**	0.048	**	-0.196	***	-0.196	***	-0.169	***
	(0.023)		(0.023)		(0.028)		(0.028)		(0.025)	
ln(selfor)	0.003		0.003		0.106	***	0.107	***	0.082	***
	(0.026)		(0.026)		(0.025)		(0.025)		(0.023)	
generality	-0.001		-0.008		1.224	***	1.179	***	1.004	***
	(0.167)		(0.167)		(0.172)		(0.173)		(0.155)	
cmade	-0.033	**	-0.036	**	-0.055	***	-0.058	***	-0.054	***
	(0.016)		(0.016)		(0.015)		(0.015)		(0.013)	
originality	0.608	***	0.624	***	-0.186		-0.159		-0.067	
	(0.141)		(0.141)		(0.160)		(0.160)		(0.143)	
ln(t)			-0.136				-0.222		-0.164	
			(0.112)				(0.110)		(0.098)	
Constant	-1.222	***	0.001		-0.976	***	1.091		0.607	
	(0.106)		(1.018)		(0.224)		(1.046)		(0.937)	
Category dummies	N		N		Y		Y		Y	
Observations	358		358		358		358		360	
Adjusted R-squared	0.27		0.27		0.58		0.58		0.59	

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: Instrumental variables estimation

	(1) lnL		(2) lnL		(3) lnLavg		(4) lnLavg	
ln(entrants)	0.091		-0.012		0.140		0.031	
	(0.129)		(0.107)		(0.113)		(0.094)	
ln(subfor)	0.086	**	-0.181	***	0.066	*	-0.157	***
	(0.039)		(0.031)		(0.034)		(0.027)	
ln(selfor)	0.032		0.178	***	0.007		0.135	***
	(0.037)		(0.048)		(0.032)		(0.042)	
generality	0.081		1.380	***	0.062		1.152	***
	(0.187)		(0.204)		(0.161)		(0.178)	
cmade	-0.045	**	-0.066	***	-0.038	**	-0.060	***
	(0.019)		(0.017)		(0.017)		(0.015)	
originality	0.631	***	-0.262		0.585	***	-0.143	
	(0.147)		(0.177)		(0.127)		(0.155)	
Constant	-1.081	***	0.006		-1.036	***	-0.095	
	(0.163)		(0.328)		(0.142)		(0.286)	
Category dummies	N		Y		N		Y	
Observations	358		358		360		360	
R-squared	0.23		0.54		0.28		0.56	

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
ln(entrants) instrumented using ln(t)

Table 6: Simultaneous equations estimation

	Lerner = L		Lerner = Lavg	
	(1)	(2)	(3)	(4)
	ln(Lerner)	ln(entrants)	ln(Lerner)	ln(entrants)
ln(entrants)	0.182***		0.188***	
	(0.036)		(0.032)	
ln(Lerner)		0.340*		0.420**
		(0.175)		(0.211)
ln(subfor)	-0.165***		-0.134***	
	(0.027)		(0.024)	
ln(selfor)	0.130***		0.105***	
	(0.026)		(0.023)	
generality	1.150***	0.389	0.941***	0.359
	(0.168)	(0.283)	(0.151)	(0.284)
cmade	-0.049***	-0.073***	-0.044***	-0.072***
	(0.014)	(0.025)	(0.013)	(0.026)
originality	-0.224	0.574**	-0.128	0.556**
	(0.156)	(0.237)	(0.139)	(0.245)
lnt		-0.097		-0.106
		(0.182)		(0.179)
offpat		1.794***		1.757***
		(0.148)		(0.157)
Constant	-0.326	0.000	-0.390	1.239
	(0.271)	0.000	(0.242)	(1.619)
Category dummies	Y	Y	Y	Y
Observations	358	358	360	360

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7: Hazard rate estimation

	Exponential distribution			Weibull distribution		
	(1)	(2)	(3)	(4)	(5)	(6)
	All	On-patent	Off-patent	All	On-patent	Off-patent
ln(entrants)	3.244***		2.811***	1.439***		2.113***
	(0.322)		(0.494)	(0.194)		(0.420)
ln(subfor)	1.898***	19.852***	1.454***	1.997***	3.789**	1.645***
	(0.186)	(12.083)	(0.173)	(0.213)	(2.274)	(0.207)
ln(selfor)	0.827*	0.328***	0.814*	0.754**	0.226*	0.778*
	(0.085)	(0.138)	(0.098)	(0.087)	(0.180)	(0.102)
generality	0.015***	0.000***	0.049***	0.012***	0.002**	0.024***
	(0.010)	(0.000)	(0.039)	(0.008)	(0.004)	(0.020)
cmade	1.143**	1.619***	1.081	1.146**	1.277	1.121
	(0.071)	(0.288)	(0.078)	(0.072)	(0.204)	(0.086)
originality	5.633**	107.233*	4.745*	11.190***	34.609*	6.557**
	(4.206)	(283.254)	(3.877)	(8.891)	(70.794)	(5.769)
Category dummies	Y	Y	Y	Y	Y	Y
Scaling parameter				6.819***	9.867***	5.415***
				0.87	2.07	1.22
Observations	2576	1277	1299	2576	1277	1299

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%