NBER WORKING PAPER SERIES

REVISITING THE PRIVATE VALUE OF SCIENTIFIC INVENTIONS

Ashish Arora
Sharon Belenzon
Elia Ferracuti
Jay Prakash Nagar

Working Paper 33056
http://www.nber.org/papers/w33056

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2024

We thank participants at the Boston University IP Day-2024 Conference for comments and feedback. We thank Hansen Zhang for providing excellent research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Revisiting the Private Value of Scientific Inventions
Ashish Arora, Sharon Belenzon, Elia Ferracuti, and Jay Prakash Nagar
NBER Working Paper No. 33056
November 2024
JEL No. O3

ABSTRACT

Estimating the private value of patents is important, yet challenging. By developing a method that uses stock market returns to produce a distribution of patent values (and not just an estimate of the mean of that distribution), Kogan, Papanikolaou, Seru, and Stoffman (2017) (KPSS) opened venues for new research. Researchers have used these estimates to compare average values of different types of patents. In this paper, we argue that KPSS values should not be used in their current form to compare mean values of different groups of patents, and show this to be the case in the context of research on the private returns to scientific patents. We extend the original KPSS method to allow for patents to be drawn from two distinct value distributions. Using this approach, we find that scientific patents have a higher mean value compared to non-scientific patents.

Ashish Arora
Fuqua School of Business
Duke University
Box 90120
Durham, NC 27708-0120
and NBER
ashish.arora@duke.edu

Sharon Belenzon
Fuqua School of Business
Duke University
100 Fuqua Drive
Durham, NC 27708
and NBER
sharon.belenzon@duke.edu

Elia Ferracuti
Fuqua School of Business
Duke University
100 Fuqua Drive
Durham, NC 27708
elia.ferracuti@duke.edu

Jay Prakash Nagar
Duke University
jayprakash.nagar@duke.edu

1 Introduction

Estimating the private economic value of patents has been a key question in the economics of innovation scholarship. Griliches (1981) was the first to estimate the returns to innovation using stock markets data. Because patent data matched to firms were not available at the time, Griliches's original approach was to estimate the stock market returns to R&D stock. The pioneering work by Hall, Jaffe, and Trajtenberg (2005) introduced firm-level patent data to the field, allowing researchers to estimate the stock market returns to patenting activity by firms. While very influential, this approach only estimates the average patent value at the firm or firm-year level, and not individual patent values, which is problematic because most patents have no or limited private (or social) value.

Kogan et al. (2017) (henceforth KPSS) made an important contribution to the field by developing a method that relies on stock market returns, as done by previous research, but also produces a distribution of patent values (and not just an estimate of the mean of that distribution at the firm or firm-year level). That is, instead of estimating the returns to patents stock, the KPSS method produces the market return to each patent separately. Estimates of the private value of patents provided by Kogan et al. (2017) have been used by researchers, typically to compare patent values across different groups of patents, such as patents that rely upon science and those that do not.

Corporate innovation increasingly relies on science (Marx & Fuegi, 2020). Science guides the creative process, reducing exploration costs, eliminating unproductive avenues, and expediting the invention timeline (Ahmadpoor & Jones, 2017; Kline & Rosenberg, 1986; Nelson, 1962; Rosenberg, 1990; Sorenson & Fleming, 2004). A recent literature attempts to compare the private value of science-based and non-science-based patents. Because scientific patents are technically more important, as measured in terms of forward citations (Arora, Belenzon, & Dionisi, 2023), and more novel, as indicated by the use of new words (Ahmadpoor & Jones, 2017; Krieger, Schnitzer, & Watzinger, 2024), than non-scientific patents, researchers have used KPSS estimates to examine whether and when the private value of scientific patents is higher than that of non-scientific patents. This research, however, has produced inconclusive and specification-sensitive evidence (e.g., Arora, Belenzon, & Dionisi, 2023; Krieger et al., 2024).

We show that KPSS values should not be used in their current form to compare the mean values of different groups of patents, such as scientific and non-scientific patents granted to a given firm. The KPSS method assumes that all patents for a given firm and year are drawn from the same distribution. Given this assumption, differences in values of different groups of patents of a firm are therefore little more than sampling variation or noise. On the other hand, if this assumption is violated, the KPSS method yields overestimates of the value of some patent groups and under-estimates of the other. Thus, KPSS estimates appear poorly suited to compare the private value of different patents based on their observed characteristics.[1] We extend the original Kogan et al. (2017) method by relaxing the assumption that the values of all patents (in a firm-year) are drawn from the same distribution to allow science-based patent values to be distributed differently from non-science-based patents.

Using this generalized approach to calculate the private value of patents, we document three important findings. First, science-based patents appear to be drawn from a different distribution than non-scientific patents. Second, forcing both types to have the same distribution (as in the original KPSS approach) underestimates the value of scientific patents and overestimates the value of non-scientific ones. Thus, the difference in value between these two groups is understated when relying on estimates produced with the KPSS method. Third, we find that this estimation process affects econometric analysis that tries to explain differences in patent values. We document that while the original KPSS estimates provide mixed and specification-dependent evidence on the science premium, our estimates produce consistently positive, economically meaningful, and statistically significant estimates of the science premium. We conclude that the mixed evidence on whether scientific patents are privately more valuable is probably because the literature has hitherto relied on KPSS esti-

[1]This is not to say that Kogan et al. (2017)'s method is flawed. What is flawed is how the resulting estimates are used to test whether patent values differ among different groups of patents.

mates, which incorrectly assume that these two types of patents are drawn from the same distribution.

Our contribution is twofold. We generalize the KPSS approach to allow for two separate distributions of patent value. Our empirical results suggest that care is required in using the KPSS approach if the intent is to explore the determinants of patent value. For instance, in addition to studying whether science-based inventions are more valuable than non-science-based patents, previous studies have used KPSS estimates to study whether larger firms have more valuable patents (Arora, Cohen, Lee, & Sebastian, 2023), the effect of anti-takeover laws on the value of patents (Keum, 2021), the effect of tax laws on patent value (Li, Ma, & Shevlin, 2021), or the effect of short-selling regulations on invention value (He, Ren, & Tian, 2023). Stated differently, many studies try to explore whether the expected value of patents depends upon a variable of interest, such as whether it is science-based, or some other characteristic of the firm (e.g., size), its environment (e.g., legal regime), or the patent (e.g., reliance on science). Our results suggest that one should estimate the distribution of patent value separately by the variable of interest, rather than use patent value estimates that assume a single distribution and then compare conditional means (e.g., by regressing patent value on the variable of interest).

Substantively, we empirically show that science-based inventions are indeed more privately valuable to firms, as inferred from stock-market responses to the patent grant. Valuable patents are more likely to be renewed to term, receive more citations from other patents in the future, and are more likely to be traded. Even controlling for value, science-based inventions receive more citations from other patents, and are more likely to be traded. However, controlling for value, science-based inventions are not more likely renewed. These patterns suggest first and foremost that the stock-market-reaction based estimate of patent value is a plausible measure of the private value of patents, and that science-based inventions are more valuable based on this measure. Second, they suggest that science-based inventions have greater spillovers for follow-on inventors. Further, science-based inventions are easier to trade, perhaps because of lower transaction cost or wider applicability, which leads to greater gains from trade.

The paper is organized as follows: Section 2 revisits the KPSS method and shows that it is poorly suited for comparing values of different groups of patents. Section 3 extends the KPSS approach to allow for such comparisons by estimating the value of different types of patents. Section 4 presents the data and our science-based patent measures. Section 5 discusses the main econometric results of the value premium that science-based patents enjoy. Empirical findings on alternate patent indicators and science premiums are presented in Section 5.4, and conclusions are summarized in Section 6.

2 Estimating the Value of Patents

2.1 Background

Estimating the private economic value of patents has been a key question in the economics of innovation scholarship. Researchers have used different approaches to measure patent value, each with its own strengths and weaknesses.

Following Sanders, Rossman, and Harris (1958), some scholars have directly surveyed inventors or firms to elicit valuations of individual patents (e.g., Gambardella, Harhoff, & Verspagen, 2008; Nagaoka & Walsh, 2009). These measures, are both subjective and sensitive to how the responses are elicited. Nonetheless, they are useful as an independent measure of value. (See also Hall and Harhoff (2012) for a more systematic review.)

A stream of literature initiated by Griliches (1981) relies on stock market value to measure the average return to the invention, namely its private value. Such estimates, however, were based on patent stocks and thus primarily at the firm or firm-year level, not at the individual patent (or patent-day) level. Kogan et al. (2017), which we describe in detail below, advanced the literature by allowing researchers to move from aggregate estimates of patent value to individual patent estimates. However, as we will show, this approach relies on assumptions that make it poorly suited to compare patent values across different groups of patents in the presence of heterogeneity across different groups of patents.

Another stream of literature measures the value of inventions using the number of forward citations received by each patent (e.g., Trajtenberg, 1990). Although patent citation trails offer valuable insights, they mainly measure technical value, namely the influence of an invention on follow-ons, which is both conceptually and empirically different from the private value of the invention (Arora, Cohen, et al., 2023). That is, patent citations are a better proxy for the societal value of the invention than for the private value of the patent itself.

A third stream of literature uses patent renewals (e.g., Harhoff, Scherer, & Vopel, 2003; Hegde & Sampat, 2009; Pakes, Simpson, Judd, & Mansfield, 1989; Schankerman, 1998; Serrano, 2010). This literature focuses on the resolution of uncertainty about the private value of either the underlying invention or perhaps of continued patent protection. Patents are treated as real options on an uncertain future profit stream. Starting from this premise, some functional-form assumptions allow researchers to derive estimates of the expected private value of a patented invention based on whether the patent is renewed or not. The Kogan et al. (2017) approach also relies on assumptions about the distribution of patent values. Consequently, as with Kogan et al. (2017), heterogeneity across different groups of patents complicates the relationship between the estimated value at the time of grant and the likelihood of subsequent renewal.

Our paper contributes to all three streams of literature. First, we extend the original Kogan et al. (2017) method by allowing different patent groups to be drawn from different value distributions, thus addressing one of the limitations in KPSS estimates. Second, our empirical results confirm that the private value of patents (computed at the time of grant) is positively associated with citations received from other patents in the future, thus suggesting a positive relation between technical and private value. Third, our empirical results are consistent with greater uncertainty in the private value of science-based patents compared to non-science-based patents, consistent with the real options framework used to estimate patent value from patent renewals.

At the same time, our methodology, which is based on the KPSS approach, shares on a limitation that extends to all three approaches to measuring patent value. It is broadly recognized that these approaches potentially confound the value of patent protection and the value of the underlying invention being protected. The value of the invention and of patent protection may also vary depending on the ability of the firm to commercialize and capture value from the invention. Specifically for our paper, we estimate whether the value of science-based inventions, as measured by whether the associated patent cites the scientific literature, is different from inventions not based on science. With this approach, we cannot distinguish whether any observable difference is because science-based inventions are more valuable, patents on science-based inventions provide more effective protection, or a combination of the two.

In addition, the values of individual patents may be interdependent. That is, there are often multiple patents covering a commercialized innovation. This makes it conceptually difficult to assign value to a given patent because it may depend on the portfolio of patents available to the firm. Finally, as reflected in the patent-renewal literature, the value of a patent evolves over time, as technology and market conditions change and as new information arrives. Therefore, the stock-market based patent values, which have built-in option value, in addition to the expected value of the patented invention itself, will tend to differ from retrospective measures that survey measures typically reflect.

2.2 The Kogan et al. (2017) approach

2.2.1 Estimation approach

Kogan et al. (2017) introduced a method that allows researchers to move from aggregate estimates of patent value to individual patent estimates. This approach relies on the idea that stock prices reflect the market's expectations of a company's future profitability, including that due to patents. The authors use the stock price reaction in a short window around patent grants to gauge their value implications for shareholders. An important advantage of this approach is that asset prices are forward-looking and hence provide us with an estimate of the private value to the patent holder that is based on ex-ante information. However, while conceptually straightforward, the approach is empirically challenging because stock prices can move for many reasons even over short periods around patent grants. KPSS overcome this difficulty by imposing distributional assumptions on the value of patent grants and other events. More specifically, KPSS's estimation approach can be summarized as follows.[2] The private value of patent j granted to firm f on day t is given by:

$$\hat{A}_{jft} = \frac{1}{K_{ft}}(1-\pi)^{-1} \cdot S \cdot \mathbb{E}[x_j|r_j] \tag{1}$$

Where K_{ft} represents the number of patents granted on the day, π is the ex-ante probability that the patent application is successful, and S denotes the market capitalization of the firm the day prior to the patent grants.[3] The key estimand is $\mathbb{E}[x_j|r_j]$, the expected value of the patent (x_j) conditional on the observed stock market reaction around the patent grant announcement date (r_j).

To estimate $\mathbb{E}[x_j|r_j]$, KPSS separate the stock return around the time patent j is granted (r_j) into two components: the change in firm value induced by the patent grant announcement (x_j) and the contemporaneous change in value induced by other value-relevant events that are unrelated to the patent (ε_j).

$$r_j = x_j + \varepsilon_j \tag{2}$$

These two components are unobservable, so the authors impose distributional assumptions to estimate them from observable variables. First, because the market value of a patent is a positive random variable, the authors assume that x_j, the stock price change due to the

[2] Please refer to Appendix A1 for a complete characterization of this approach.

[3] To reduce notation, we omit the length of the time window over which the stock market response to the patent grant is measured.

patent grant, follows a normal distribution truncated at zero: $x_j \sim \mathcal{N}^+(0, \sigma^2_{xft})$.[4] Second, because value-relevant events other than patent grants can be both value-enhancing (e.g., new product announcement) or value-destroying (e.g., lawsuit announcement), ε is assumed to be normally distributed: $\varepsilon_j \sim \mathcal{N}(0, \sigma^2_{\varepsilon ft})$. Under these assumptions, $\mathbb{E}[x_j|r_j]$ can be estimated as follows:

$$\mathbb{E}[x_{jft}|r_{jft}] = \delta_{ft} r_{jft} + \sqrt{\delta_{ft}}\sigma_{\varepsilon ft}\frac{\phi(R_j)}{1-\Phi(R_j)} \tag{3}$$

where $R_j = -\sqrt{\delta_{ft}}\frac{r_j}{\sigma_{\varepsilon ft}}$, $\delta_{ft} = \frac{\sigma^2_{xft}}{\sigma^2_{\varepsilon ft}+\sigma^2_{xft}}$.

Equation 3 has two attractive features: (i) the estimated patent value is higher in the presence of higher stock returns at grant date (r_{jft}); and (ii) the relation between estimated patent value and stock returns at grant date is stronger when stock markets are more informative about patent value, namely they have less noise (i.e., larger δ_{ft}).

Implementing this formula requires researchers to estimate two parameters: δ_{ft} (the signal-to-noise ratio) and $\sigma^2_{\varepsilon ft}$ (the variance of returns due to other contemporaneous unobservable events, also called the noise). KPSS assume that the signal-to-noise ratio is constant over time and across firms, so that $\delta_{ft} = \delta$, to economize on the number of parameters to be estimated, and estimate it using the following regression:

$$ln(r_{fd})^2 = a_0 + a_{ft} + b_d + \gamma I_{fd} + \mu_{fd} \tag{4}$$

Where r_{fd} denotes the three-day idiosyncratic return for firm f starting on the day d, I_{fd} is an indicator that identifies days with patent grants, and a_{ft} and b_d indicate controls for firm-year and the day of the week fixed effects, respectively, to account for the time-varying nature of firm volatility and its seasonal fluctuations. The coefficient of interest is γ, which denotes the effects of a patent grant announcement on idiosyncratic stock return volatility and can be used to recover $\hat{\delta} = 1 - e^{-\hat{\gamma}}$. Finally, the authors estimate $\sigma_{\varepsilon ft}$ non-parametrically

[4] This is the same as assuming that granted patents at a minimum do not destroy firm value.

using the sum of squared market-adjusted returns.

It is worth noting that as equation 3 indicates, KPSS deal with the presence of multiple patents granted to a firm on the same day by first estimating the total value attributable to patent grant on a given day ($\mathbb{E}[x_j|r_j]$), and then by dividing it equally among all patents granted on that day. This is equivalent to assuming that all patents issued to a firm on a given day are identical in value. Alternatively, one could assume that all patents issued are identically (and independently) distributed. This modification would require an adjustment to account for the number of patents granted on a particular day when estimating the signal-to-noise ratio. As described in appendix section A2.1, the KPSS method tends to overestimate patent value compared to the average when the number of patents granted on that day is low, while adjusting for the number of patents underestimates the patent value compared to the average in the same scenario. Although the two methods show little difference in predicting technical quality (forward citations) within firms, the modified approximation performs worse when firm-fixed effects are excluded. Accordingly, in the paper we stick with the baseline KPSS method due to its simplicity.

2.2.2 KPSS values and underlying patent value

Note that the KPSS values are distributed differently than the distribution of x itself. Formally, the KPSS value is a random variable whose distribution depends on the distribution of the market return, of which the value of the patent is but one component. Numerical simulations detailed in Appendix A indicate that the distribution of values obtained using the KPSS approach has the same mean as the underlying distribution of patent value, but has a smaller variance and is less skewed. This should be kept in mind when comparing patent values obtained using the KPSS approach with those from surveys or patent-renewal estimates.

2.2.3 Using KPSS to compare mean values of groups of patents

In light of the popularity of the KPSS measure as the private value of patents, it is important to investigate and understand its strengths and limitations. Our paper takes a step in this direction in the context of research that compares the economic value of patents based on their observed characteristics. We argue that comparing the means of different patent types ex-post, namely after having implemented the KPSS method described above, is problematic for both conceptual and empirical reasons.

First, comparing the means of different patent types ex-post, namely after having implemented the KPSS method described above, is conceptually problematic because inconsistent with a key estimation assumption. KPSS assume a common signal-to-noise ratio (δ) to all patent types granted to the focal firm in a given year, which is equivalent to assuming that all patent *groups* granted to focal firm in a given year are drawn from the same distribution.[5] In turn, this implies that the returns for all patent groups for a given firm in a given year are drawn from the same distribution. Therefore, the expected value of different patent groups is the same, and any observable difference between their sample means should reflect noise.

To clarify, suppose there are two patent groups, S and N respectively. The focal firm receives S patents of type S and N patents of type N during the year. The KPSS estimates of the change in firm value induced by the i^{th} S type patent is $\mathbb{E}[x_{is}|r_{is}] = \delta r_{is} + \sqrt{\delta}\sigma_\varepsilon \lambda(R_{is})$ where $\lambda(R_{is}) = \dfrac{\phi(R_{is})}{1-\Phi(R_{is})}$ and $R_{is} = -r_{is}\dfrac{\sqrt{\delta_s}}{\sigma_\varepsilon}$. The sample mean of all S type patents (for that firm and year) using KPSS value, $\frac{1}{S}\sum_{i=1}^{i=S}\mathbb{E}(x_{is}|r_{is})$ denoted by $\widehat{\mathbb{E}[x_s|r_s]}$, is:[6]

$$\widehat{\mathbb{E}[x_s|r_s]} = \delta \frac{1}{S}\sum_{i=1}^{i=S} r_{is} + \sqrt{\delta}\sigma_\varepsilon \frac{1}{S}\sum_{i=1}^{i=S}\lambda(R_{is}) \tag{5}$$

Therefore, the sample difference between the mean value patent groups S and N is:

[5] Other assumptions are that (i) x_j follows a truncated normal distribution at zero, denoted as $\mathcal{N}^+(0, \sigma^2_{xft})$; (ii) ε_j follows a normal distribution, expressed as $\mathcal{N}(0, \sigma^2_{\varepsilon ft})$; and (iii) the ex-ante probability of a successful patent grant does not vary with the economic value of a patent.

[6] In large samples, the sample mean converges in probability to $\mathbb{E}_{r_s}\mathbb{E}(x_s|r_s) = \mathbb{E}(x)$

$$\widehat{\mathbb{E}[x_s|r_s]}-\widehat{\mathbb{E}[x_n|r_n]} = \delta\left(\frac{1}{S}\sum_{i=1}^{i=S} r_{is} - \frac{1}{N}\sum_{i=1}^{i=N} r_{in}\right)+\sqrt{\delta}\sigma_\varepsilon\left(\frac{1}{S}\sum_{i=1}^{i=S}\lambda(R_{is}) - \frac{1}{N}\sum_{i=1}^{i=N}\lambda(R_{in})\right) \quad (6)$$

Each bracketed term reflects the difference in the means of two samples drawn from the same distribution. The market return for a given firm (for patent grant days) is $r_k = x_k + \varepsilon, k \in \{S, N\}$. By assumption, x_s and x_n have the same distribution. This implies that r_s and r_n have the same distribution (as do $\lambda(R_{is})$ & $\lambda(R_{in})$), so for large samples, the bracketed terms will vanish. As a consequence, *under the KPSS assumptions there can be no systematic differences in the means of value for any groups of patents for a given firm and year.*

Second, comparing the means of different patent types ex-post, namely after having implemented the KPSS method described above, is empirically problematic because it may mis-estimate the true difference in value between different patent groups. Suppose that S type patents have a higher value than N type patents, i.e., $\sigma_s > \sigma_n$, but the researcher follows the KPSS approach. Because this approach imposes a single parameter δ_x, the researcher is implicitly imposing a common σ_x for all the patents granted to the firm in the year that is a weighted mean of the true parameters i.e., $\delta_s \geq \delta_x \geq \delta_n$. Under these conditions, the sample mean for S type patents using patent values computed from KPSS will underestimate the true value of S type patents, and conversely, the sample mean for N type patents will overestimate the true value of N type patents.

To see this, note that $r_{is} = x_{is} + \epsilon_i$, and $r_{jn} = x_{jn} + \epsilon_j$. If ε_i and ε_j have a normal distribution with mean zero and standard deviation σ_ε, then letting r_s represent the market returns for S type patents and r_n for N patents, we have

$$\mathbb{E}[r_s] = \mathbb{E}[x_s] = \sigma_s\sqrt{\frac{2}{\pi}} > \mathbb{E}[r_n] = \mathbb{E}[x_n] = \sigma_n\sqrt{\frac{2}{\pi}} \quad (7)$$

The other term in equation 5 involves $\lambda(R)$. Taking expectations, equation 3 implies that

$$\mathbb{E}[\lambda(R_s)] = \mathbb{E}[r_s]\frac{(1-\delta_s)}{\sqrt{\delta_s}\sigma_\varepsilon}; \quad \mathbb{E}[\lambda(R_n)] = \mathbb{E}[r_n]\frac{(1-\delta_n)}{\sqrt{\delta_n}\sigma_\varepsilon} \tag{8}$$

Using equation 5 one gets

$$\begin{aligned}
\widehat{\mathbb{E}[x_s|r_s]} &= \delta\frac{1}{S}\sum_{i-1}^{i=S} r_{is} + \sqrt{\delta}\sigma_\varepsilon\frac{1}{S}\sum_{i-1}^{i=S}\lambda\left(R_s\sqrt{\frac{\delta}{\delta_s}}\right) \\
&= \delta\,\mathbb{E}[r_s] + \sqrt{\delta}\sigma_\varepsilon\,\mathbb{E}\left[\lambda\left(R_s\sqrt{\frac{\delta}{\delta_s}}\right)\right] \text{ (when } S\to\infty) \\
&\le \delta\,\mathbb{E}[r_s] + \sqrt{\delta}\sigma_\varepsilon\,\mathbb{E}[\lambda\,(R_s)] \text{ (because } \lambda \text{ decreases with } R \text{ and } \delta_s > \delta) \qquad (9) \\
&= \delta\,\mathbb{E}[r_s] + (1-\delta_s)\,\mathbb{E}[r_s]\text{(by using equation 8)} \\
&\le \mathbb{E}[r_s] \text{ because } \delta_s \ge \delta \\
\implies \widehat{\mathbb{E}[x_s|r_s]} &\le \mathbb{E}[x_s]
\end{aligned}$$

That is, the sample mean for S type patents using KPSS values will underestimate their true mean. A similar logic implies that the sample mean for N type patents using KPSS values with over-estimate their true value.

In sum, if researchers are interested in testing for differences in the mean value of two groups of patents, comparing the means of different patent types ex-post after having implemented the KPSS method is conceptually inconsistent. Furthermore, if one does find systematic differences between types of patents (within a firm-year) in patent value, the estimated differences will understate the true differences. Finally, care is needed when comparing patent values across firms. Even when two firms have the same distribution for δ, the patent values also depend on σ_ε, which may differ across firms.

3 Estimating the Value of Patents with Multiple Patent Types

We extend the KPSS estimation method to (i) allow the value of different patent types to be drawn from different distributions and (ii) incorporate the possibility that the two patent groups are granted to the focal firm on the same day. Consider two patent types (for example, scientific and non-scientific patents) denoted by S and N. We assume that the values of these patents, x_{sj} and x_{nj} are distributed according to a normal distribution truncated at zero as in KPSS but with different variances (σ^2_{xsft} and σ^2_{xnft}). Assume additionally (as in KPSS) that the value of other contemporaneous unobservable events is normally distributed with variance $\sigma_{\varepsilon ft}$, and that the three variances may vary across firms and over time, but in constant proportions.

3.1 Two patent types granted on different days

If S and N patents are granted on separate days, we can simply extend the KPSS approach by estimating patent type-specific δs. To do this, it suffices to modify equation (4) as follows:

$$ln(r_{fd})^2 = a_0 + a_{ft} + b_d + \gamma_s I_{sfd} + \gamma_n I_{nfd} + \nu_{fd} \tag{10}$$

where I_{sfd} indicates announcement days when S type patents are granted, while I_{nfd} identifies announcement days where N type patents are granted. Comparing equation (10) with equation (4), we see that KPSS implicitly assumes that $\gamma_s = \gamma_n = \gamma$. It follows that in KPSS γ is a weighted average of γ_s and γ_n. So, if S type patents have a higher signal-to-noise ratio, then $\gamma_n \leq \gamma \leq \gamma_s$. It follows that a higher γ_s (γ_n) implies a higher σ_{xsft} (σ_{xnft}), and therefore a higher mean value of S (N) patents, because $\sigma_{\varepsilon ft}$ is effectively assumed to be constant for a given firm year.

Thus, estimating a single δ as in equation 4, which forces the signal-to-noise ratio to be the same for all patents, results in an underestimation of the value of S type patents and an overestimation of the value of N type patents, as formally shown earlier in section 2.2.3. Estimating separate, patent type-specific δs via equation (10) overcomes this limitation and allows researchers to recover the correct patent value for the two different patent types.

3.2 Two patent types granted on same day

The modification detailed in the previous subsection is insufficient when both patent types can be granted on the same day. This is because when different patent types are granted on the same day, the observed stock market response on that day arises from two separate patent values and two different signal-to-noise ratios. As a consequence, equation 3 is not longer valid. We adopt the method from Papadopoulos (2015) and Papadopoulos (2021) and derive the adjusted formula for the value of S-type patents when different patent types are granted on the same day:[7]

$$E(x_{sj}|r_{sj}) = \frac{\theta_1^2}{(1+\theta_1^2+\theta_2^2)} r_j + \frac{\left\{2\frac{(1+\theta_2^2)\sigma_{\varepsilon ft}^2}{\omega_1}\phi(r_j/\omega_1)\Phi\left(\frac{\lambda_1}{\omega_1}r_j\right) - \frac{2\theta_1^2\sigma_{\varepsilon ft}^2}{\omega_2}\phi\left(\frac{r_j}{\omega_2}\right)\Phi\left(\frac{\lambda_2}{\omega_2}r_j\right)\right\}}{\left[\Phi\left(\frac{r_j}{\omega_1}\right) - 2T\left(\frac{r_j}{\omega_1},\lambda_1\right) + \Phi\left(\frac{r_j}{\omega_2}\right) - 2T\left(\frac{r_j}{\omega_2},\lambda_2\right)\right]} \tag{11}$$

Where:

- $\theta_1 = \frac{\sigma_{xsft}}{\sigma_{\varepsilon ft}}$ and $\theta_2 = \frac{\sigma_{xnft}}{\sigma_{\varepsilon ft}}$
- $\omega_1 = \frac{s\sqrt{1+\theta_2^2}}{\theta_1}$ and $\omega_2 = \frac{s\sqrt{1+\theta_1^2}}{\theta_2}$
- $\lambda_1 = \frac{\theta_2}{\theta_1}\sqrt{1+\theta_1^2+\theta_2^2}$ and $\lambda_2 = \frac{\theta_1}{\theta_2}\sqrt{1+\theta_1^2+\theta_2^2}$
- $s = \sqrt{\sigma_{\varepsilon ft}^2 + (\sigma_{xsft})^2 + (\sigma_{xnft})^2} = \sigma_{\varepsilon j}\sqrt{1+\theta_1^2+\theta_2^2}$
- T = Owen's T function (Owen, 1980)

[7] Please refer to Appendix B1 for a step-by-step derivation.

Equation 11 is a generalized version of equation 3 from KPSS: when only S-type patents are granted on a given patent day, $\sigma_{xnft} \to 0$ so $\mathbb{E}[x_{sj}|r_{sj})$ reduces to the equation (3). Equation 11 accommodates the presence of two signals on the same day, with both S and N type patents, and these signals are distributed differently, albeit independently.

As in KPSS, our approach requires empirical counterparts for parameters related to signal-to-noise ratios, δ_s and δ_n, and the noise variance, $\sigma_{\varepsilon ft}$. To recover the signal-to-noise ratios we estimate the following equation:

$$ln(r_{fd})^2 = a_0 + a_{ft} + b_d + \gamma_s I_{sfd} + \gamma_n I_{nfd} + \nu_{fd} \tag{12}$$

where r_{fd} denotes the three-day idiosyncratic return of firm f following the patent grant date d. The equation includes separate dummy variables for different patent types. Specifically, I_{sfd} indicates announcement days that include science-based patents, and it is defined as a dummy variable set to one if at least one science-based patent is granted on that day, zero otherwise; while I_{nfd} identifies announcement days that do not include science-based patents, and it is defined as a dummy variable set to one if at least one non-science-based patent but no science-based patent is granted on that day, zero otherwise. As before, we use the estimated values of γ_s and γ_n to recover estimates of δ_s and δ_n respectively.

We calculate $\sigma_{\varepsilon ft}$ as in KPSS with:

$$\sigma^2_{\varepsilon ft} + \mu_{sft}\sigma^2_{xsft}(1+l) + \mu_{nft}\sigma^2_{xnft}(1+l) = \sigma^2_{ft}(1+l) \tag{13}$$

Here, μ_{sft} represents the share of days when science patents are granted and μ_{nft} represents the share of days when only non-science-based patents are granted in a given firm-year. σ^2_{ft} is the estimated realized mean idiosyncratic squared returns of firm f at time t, and l is the window of time in days over which the market reaction is calculated.[8]

[8] The term $1+l$ represents the time window (in days) over which the stock market's reaction is calculated, indicating how long it takes for the stock market to incorporate information about the patent's value into stock prices. In contrast, noise is continuously integrated into stock prices throughout the year.

Table 1: Overview of patent value estimation methods

Patent value estimates	Description
Patent Value: KPSS	The private value of a patent is estimated using the KPSS methodology described in Section 2.2.1. This method estimates a common signal-to-noise ratio, δ, assuming that all patents for a focal firm in a given year are drawn from the same distribution.
Patent Value: Sci & Non-Sci (separate)	The private value of patents is estimated using distinct distribution assumptions for science-based and non-science-based patents. We calculate the signal-to-noise ratio for each patent type (δ_s & δ_n) using Equation 12. However, on days when at least one science-based patent is granted, the patent value is distributed equally across all patents, as described in Section 3.
Patent Value: Sci & Non-Sci (mixed)	The patent's private value, estimated using separate distribution assumptions for science-based and other patents. we use the method described in Section 3.2 to estimate the patent private value on science days when both types of patents are present (at least one science-based and one non-science-based patent).

3.3 Simulations

We run a simulation analysis to illustrate the differences between KPSS and our procedure. We simulate 5,000 firms, and for each firm, we generate daily data for 40 years, or 365×40 days. For each firm i, we draw σ_{ni} values from a normal distribution with a mean of 0 and a standard deviation defined as a random number between 0 and 1, denoted as σ_i. The values of x_n are drawn from $\mathcal{N}^+(0, \sigma_i)$. For S-type patents, we assume $x_s \sim \mathcal{N}^+(0, 1.1 * \sigma_i)$ ($\sigma_{si} = 1.1\sigma_i$). The noise distribution follows $\epsilon \sim \mathcal{N}(0, 5 * \sigma_i)$ ($\sigma_{\epsilon i} = 5\sigma_i$). For 15% of the days, x_n is drawn from a half-normal distribution $\mathcal{N}^+(0, \sigma_i)$, representing N-type patents and for 10% days x_s is drawn from a half-normal distribution $\mathcal{N}^+(0, 1.1 * \sigma_i)$. Next, we calculate the realized value $r = I(n) \cdot x^n + I(s) \cdot x^S + \epsilon$, where $I(n)$ and $I(s)$ are indicator functions for the N-type and S-type patents, respectively, x_n and x_s are the corresponding values, and ϵ is the noise term. After generating the data, we first compute $\mathbb{E}[x_j|r_j]$ following the KPSS methodology. To estimate the common signal-to-noise ratio we estimate equation 4 for our simulated data. We also estimate separate signal-to-noise ratios using equation 10, and compute $\mathbb{E}[x_j|r_j]$ separately for S-type patents and N-type patents.

The simulation illustrates that when we do not account for different signal-to-noise ratios for each patent type, we significantly underestimate the average difference between the values of the two groups. In our simulated data, the actual difference in the average between the two types of patents, $\mathbb{E}[x_s]$ and $\mathbb{E}[x_n]$, is 0.04 (10%). When we impose a common signal-to-noise ratio (as in KPSS), the average difference is very close to 0. However, when we allow the signal-to-noise ratio to vary between S-type and N-type patents, the difference between $\mathbb{E}[x_s]$ and $\mathbb{E}[x_n]$ increases to 0.0447 (10.8%)[9]

4 Private Value of Science and non-Science Patents

4.1 Data and Measurement

We assemble a dataset of U.S. Patent and Trademark Office (USPTO) patents assigned to U.S.-based firms and between 1980 and 2019, where we restrict our sample to 2019 to avoid COVID-19 disruptions. This dataset is constructed by merging various sources. The starting point is the DISCERN 2.0 dataset, which connects all USPTO patents granted to Compustat firms (Arora, Belenzon, & Sheer, 2021a) and provides significant enhancements and expansions to the historical NBER patent dataset. We complement this database by gathering data on the patent filing date, International Patent Classification (IPC) classes, patent applicants, inventors, and other relevant details from the USPTO patent database, and forward citations received by a patent, the number of claims, family size and other related indicators from the OECD patent quality indicators database (Squicciarini, Dernis, & Criscuolo, 2013).

We link the resulting database with the "Reliance on Science" dataset, which includes all patent citations to scientific publications (Marx & Fuegi, 2020),[10] which assigns a confidence

[9] In the estimation equation 4, the coefficient of $\hat{\gamma}$ is 0.042. However, when we allow for two signal-to-noise ratios, the estimated coefficients are $\hat{\gamma_n} = 0.047$ and $\hat{\gamma_n} = 0.038$. The actual value of the S patent dummy coefficient in our simulated data is $\ln\left(\frac{(1.1)^2+5^2}{5^2}\right) = 0.47$ and for the N patent dummy coefficient is $\ln\left(\frac{1^2+5^2}{5^2}\right) = 0.39$, as shown in Appendix Table D1. A detail description and results are provided in Appendix Table D1-D4

[10] https://zenodo.org/records/8278104

score to each citation between a patent and a paper. This score measures the accuracy of the linkage in reflecting a true citation. We include all patent-publication pairs with a confidence score of 5 or higher, corresponding to a precision of 99.53% and a recall of 92.02%. We complement this data by sourcing information on the scientific articles, including publication year, journal details and publication-publication citations, from Microsoft Academic (Sinha et al., 2015), a large-scale scholarly database, and OpenAlex (Priem, Piwowar, & Orr, 2022), an open-access dataset that builds on the Microsoft Academic Graph (MAG) and contains scholarly publications such as journal articles, conference papers, books and book chapters. Finally, we incorporate stock market and financial data on firms from CRSP and Compustat.

Our analysis sample consists of the set of patents granted by the United States Patent and Trademark Office (USPTO) from 1980 to 2019. To focus on the innovating firms, we restrict the analysis to a subset of 4,977 firms that were assigned at least one patent during the sample period. Our final sample comprises 1,448,025 patents, all assigned to publicly traded firms headquartered in the United States.

4.1.1 Classifying Patents in Science-based and Non-Science-based

We classify patents as science-based and non-science-based by relying on non-patent literature (NPL) citations in patent documents. While this approach is imperfect and has known limitations (Cotropia, Lemley, & Sampat, 2013; Narin, Hamilton, & Olivastro, 1997; Roach & Cohen, 2013), non-patent literature citations represent a direct and observable link between invention and science that we can exploit for our classification.

Over the years, both the total number of patents and the proportion of patents citing at least one Non-Patent Literature (NPL) have experienced significant growth. In our sample, the percentage of patents citing at least one NPL rose from 19.4% in 1980 to 41.6% in 2019. Furthermore, among patents citing NPL, the average number of NPL citations per patent increased from 2.43 in 1980 to 14.83 in 2019.[11] There is also substantial variation in citing

[11]Our sample focuses exclusively on public firms and includes both in-text and front-page citations. In this cohort, patents issued in 1980 averaged fewer than one citation to scientific literature, while more recent

patterns across patent classes, with biotechnology and other life science fields leading in terms of the number of papers cited in a patent (Branstetter & Ogura, 2005).

Accordingly, we define "science-based patents" as those in the top three quartiles for the number of NPL citations within a given IPC class (at 4 digit level) and year, provided they cite at least one NPL. To ensure the robustness of our results, we explore an alternative definition by categorizing science-based patents as those with above-median NPL citations, still contingent on citing at least one NPL within a given IPC class and year.

4.2 The Private Value of Science and Non-Science Patents

We estimate the private value of science-based and non-science-based patents applying three different estimation methods to this sample. First, we follow the KPSS approach as described in Section 2.2.1. Second, we estimate two signal-to-noise ratios as outlined in Section 3 and divide the patent value equally across all patents on days when at least one science-based patent is granted, and likewise for patent days with non-science patents. Third, we again estimate two signal-to-noise ratios as outlined in Section 3 but this time we apply our methodology from 3.2 to estimate the private patent value when both types of patents (science-based and non-science-based) can be granted on the same day. We refer to these three estimates as *KPSS*, *Sci&Non-Sci (separate)*, and *Sci&Non-Sci (mixed)*, as shown in Table 1.

For the baseline KPSS computations, we estimate the (common) signal-to-noise ratio from equation 4, which produces an estimated value of $\hat{\gamma}$ of 0.0192, similar to the value of 0.0145 estimated in Kogan et al. (2017). We use this parameter value in conjunction with equations 1 and 3 to estimate the private value of patents. Table 2 shows that the mean patent value is \$17.516 million and the median value is \$7.435 million with a standard deviation of 42.109 (all amounts in 1982 constant USD). These values are higher than the estimates in Kogan

patents have increased that average to over six citations each. Academic patents have much higher citation (only front-page citations) rates for scientific work, averaging 14 citations, while corporate patents (from both public and private firms) average just 2, and government patents average 1.3 (Marx & Fuegi, 2020).

et al. (2017) (mean = \$10.36 million; median = \$3.22 million; SD = \$32.04 million), possibly because our sample period is shorter and more recent than theirs (1980-2019 as opposed to 1926-2010), and because we use DISCERN, which matches patents to Compustat firms with greater accuracy compared to the NBER patent file (Arora, Belenzon, & Sheer, 2021b). When we separate patents in science-based and non-science-based, we find that scientific patents are more valuable: their average dollar value is \$21.007 million, as compared to an average dollar value of \$16.413 million for non-scientific patents (again all amounts are in 1982 constant USD).

Next, we use *Sci&Non-Sci (separate)*, which entails the following procedure. First, we estimate the signal-to-noise ratio for each type of patent from Equation 12 and patent value estimates as described in 3. Next, we estimate the patent private value for two types of patent days: days when at least one science-based patent is granted and days when only non-science-based patents are granted. Similar to KPSS, we distribute the private value equally among all patents that were granted on the science patent day, even if they are not science-based patents. We find that the estimated value of $\hat{\gamma_s}$ is 0.0224, while the estimated value of $\hat{\gamma_n}$ is 0.0174. Put differently, scientific patents explain more of the variation in stock returns than non-scientific patents, implying higher estimated values relative to the KPSS estimates.Consistent with this intuition, we find that using this procedure, the average difference in private value between science-based and non-science-based patents increases to 5.910 million USD, compared to 4.594 million USD in the baseline KPSS.

Last, we estimate patent value using the method *Sci&Non-Sci (mixed)* described in Section 3.2, which accounts for the possibility that both types of patents are granted on the same day. More specifically, we use the signal-to-noise ratios for science-based and non-science-based patents estimated using Equation 12 in conjunction with equations 1 and 11 to compute the private value of patents. We find in Table 2 that the average patent value is \$17.232 million and the median is \$7.230 million, with a standard deviation of \$41.783 million. Importantly, when we separate patents in science-based and non-science-based, we find that the average value of science-based patents \$22.563 million, while the average dollar value of non-science-based patents is \$15.546 million, as shown in Table 3. Thus, the KPSS methodology underestimates the value of scientific patents by an average of 7.5%, while overestimating the value of other patents by 5.5%, as expected.[12]

4.3 Validation: the private and scientific value of invention

The literature has extensively investigated the relation between the private and scientific value of inventions, showing that forward citations (citations received from other patents) are a useful proxy of invention quality and are linked to higher private and social patent values(Hall et al., 2005; Kogan et al., 2017; Nicholas, 2008).

Kogan et al. (2017) show that their patent value estimate is related to future citations, and interpret this evidence as validating the external validity for their measure. In Table 4, we reexamine the relationship between patent forward citations and both the KPSS and our measures of the market value of patents. Our main dependent variable is the patent forward citations received over a five-year period. We also define a dummy variable that equals 1 if the patent is among the top 1% most cited patents up to 5 years after publication within a given year and technical class.

Column 1-4 Table 4 show that the percentage of variation in future citations explained by the private value of the patent increases when we measure this value with our adjusted approach instead of relying on KPSS estimates. The regression coefficient for private value increases from 0.083 in the baseline KPSS to 0.088 for the patent value estimate *Sci&Non-Sci (separate)*, and further to 0.095 for patent value estimates *Sci&Non-Sci (mixed)*, reflecting a 15.30% percentage increase in coefficient compared to baseline KPSS. We find similar results for the breakthrough patent dummy variable that takes value 1 if the patent is in the top

[12] Recall that if all patents for a firm-year were drawn from the same distribution, the average difference between science and non-science patents for firm-year would tend to vanish. The average difference in the value of science-based patents and other patents (over firms and years) is \$7 million using our method, whereas it is \$4.58 million in the KPSS estimate. The positive KPSS estimate is inconsistent with the assumption that science and non-science patents are drawn from the same distribution for a given firm-year.

1% of cited patents, as shown in Column 5-8 Table 4.

Overall, this analysis indicates that the link between the private and scientific value of invention becomes stronger with our estimates of patent value. Allowing for the value of science-based patents to differ from non-science-based patents appears to produce estimates of patent value that are more strongly related to future citations received by the patent.

5 Patent private value and science premium

5.1 Main analysis

The extant literature has used KPSS estimates to study whether/how the mean value of patents varies by patent type (e.g., science-based or not). We have discussed that doing so is inconsistent with the assumptions embedded in the KPSS estimates, which assume that the values for all patents (for a given firm and year) are drawn from the same distribution. In particular, these assumptions imply that within a firm, variations in the average value of groups of patents in a given period must be random. We have also shown that doing so is likely to bias the difference in mean value between different patent types toward zero when the two patents are drawn from different distributions. Indeed, the estimated within-firm differences in value between science-based and non-science-based patents tend to be small and unstable (Krieger et al., 2024). Our generalization does not force this homogeneity and therefore potentially allows for differences to exist even within a firm-year. We empirically examine this possibility in this section.

We follow Krieger et al. (2024) and assume that the value of a patent consists of an additive separable technological component, a proximity to science component, an idiosyncratic component, and a time component. Accordingly, we use the following regression specification:

$$Y_{ijkt} = \beta_0 + \beta_1(\text{Science dummy}_i) + \eta X_i + \alpha_t + \delta_j + \psi_k + \epsilon_i \tag{14}$$

where Y_{ijt} is the log transformation of the dollar value of patent i with (IPC 4-digit) technology class j, patent grant year t, and firm k. We measure Y_{ijt} using the KPSS approach first, and our extentions, which relax the homogeneous distribution assumption, second. X_i are measures patent characteristics including technical quality (measured by forward citations within a five-year period).

We recognize that including controls for patent characteristics such as technical quality is inconsistent with the model used to generate estimates of patent value. However, we do so to facilitate comparison with the literature. The coefficient β_1 represents the *science premium*, which the foregoing analysis suggests will be underestimated using KPSS-generated patent values.

We estimate equation 14 on our sample, and report estimated coefficients in Table 5. We present estimates using the dollar value of patents based on the KPSS approach in Columns 1-3. In Column 1, in which we do not include firm fixed effects and Patent IPC class fixed effects, the estimated KPSS dollar value is 11.98% higher for science-based patents compared to non-science-based ones, indicating the presence of a science premium.[13] The inference changes dramatically when we incrementally include patent technical field (Column 2) and firm fixed (Column 3), with the difference in the KPSS patent value approaching zero as our specification becomes more stringent. Overall, we conclude that the KPSS estimates provide evidence of the *science premium* that is mixed and sensitive to the specification.

Next, we estimate the same specification using patent value estimates calculated with our methodology, where the signal-to-noise ratios are estimated according to equation 10. In Columns 4-6 of Table 5 we present results using the dollar value of patents estimated with the *Sci&Non-Sci (separate)* approach, which assumes separate distributions for science-based and non-science-based patents but equally assigns the patent value across all patents granted on that day. Column 4 reports the estimated coefficients without patent class or firm fixed effects, yielding a coefficient estimate of 0.1593—higher than the 0.1132 estimate using the

[13] We calculate this amount as $(e^{0.1132} - 1) \times 100$.

KPSS approach discussed earlier and thus consistent with KPSS values underestimating the *science premium*. In Columns 5 and 6, we incrementally include IPC fixed effects and firm fixed effects, respectively. The estimated coefficient decreases to 0.0423 (4.3%) when both patent technical and firm fixed effects are included.

In Columns 7-9, we present results using the dollar value of patents estimated with the *Sci&Non-Sci (mixed)* approach, which assumes separate distributions for science-based and non-science-based patents and relies on equation 12 to separate the value of scientific and non-scientific patents granted on the same day. The table show that with our most conservative estimate, which includes IPC fixed effects and firm fixed effects, the science premium increases to 14.27%, compared to 4.3% when assuming separate distributions for science-based and non-science-based patents but equally dividing the value among all patents granted on science patent days (defined as days when at least one science-based patent is granted) and only 0.08% when the value of different patent types is assumed to be drawn from the same distribution as in KPSS.

These findings are consistent with the view that the existing empirical evidence on the science premium is mixed because KPSS estimates are poorly suited to compare the private value of patents of different types, as in the case of science-based and non-science-based patents.

5.2 The economic value of patents by technological sector

In additional analyses, we use the International Patent Classification (IPC) system, which includes 35 subgroups as defined by the World Intellectual Property Organization (WIPO), to classify patents in the 4 fields - Life Sciences, Chemicals, Information and Communication Technology (ICT), and others. We then re-estimate the baseline regression equation separately for observations in each technological class. Table 7 shows that the lowest *science premium* is observed in Life Sciences (10.34%), followed by Chemicals (12.92%), Information and Communication Technology (ICT) (13.94%), and all other remaining sectors (15.04%).[14] In Appendix C, we perform a similar analysis using KPSS values instead, and find different results. Using the KPSS measure, the science premium exists only in ICT and Other sectors (where it is economically small), but not in Life Science or Chemicals. We interpret this as further confirmation that the using KPSS estimates based on a single distribution produces biased estimates of the patent premium.

5.3 Robustness checks

Our results continue to hold when using alternative definitions of science dependency, such as classifying science-based patents as those with an above-median number of NPL citations, conditional on citing at least one NPL within a given IPC class and year. Column 2 of Table 6 displays the difference in the dollar value of the private value of patents. The estimated coefficient for the science dummy increases to 0.1654 from 0.1334, as expected, reflecting the narrower definition of science.[15]

In our baseline analysis, we include all patent-publication pairs with a confidence score of 5 or higher. We replicate our analysis with a higher minimum confidence score of 9, which results in a slight loss in recall (fewer science-based patents) but improves the precision ($\approx$ 99.97% in identifying patent citations to scientific literature). The estimated coefficient of the science dummy is 0.1608, compared to 0.1334 for the baseline analysis (confidence score of 5 or higher). The increase in the estimated value of the *science premium* aligns with expectations, driven by the higher precision in capturing the science dependency of patents.

5.4 Science-based inventions and patent quality

Previous studies have advanced multiple explanations for why there exists a "science premium". For instance, (Krieger et al., 2024) attribute this premium to patent novelty, as

[14] Appendix Table C1 provide the results where patent value is equally distributed among all patents of that type granted on the day.

[15] We re-estimate equation 4 for γ_S and γ_N as shown in Table 6.

measured by the combination of new words in patent documents. If more novel patents are more valuable, and science-based patents are more novel, a science premium would be observed. As discussed earlier, the assumptions underlying the KPSS method and our generalization thereof are inconsistent with regressing patent value on patent characteristics such as novelty. Instead, in this section, we empirically explore the link between science and indicators of patent quality, controlling for patent value, estimated separately for science-based and non-science-based inventions. All specifications include controls for year, IPC, and firm fixed effects.

Forward citations received:: Follow-on citations are a common measure of patent quality (e.g., Ahmadpoor & Jones, 2017; Sorenson & Fleming, 2004; Trajtenberg, 1990). In Column 1 of Table 8, we confirm that there exists a positive association between science and forward citations, a commonly used measure of science premium. Moreover, this association is virtually unchanged after we control for patent value, as shown in Column 2. The implication is that science-based patents receive more citations in part because they represent higher quality (more valuable) inventions, but also because they are more valuable as inputs into follow-on inventions. Put differently, a science-based invention creates more spillovers than a non-science-based invention that is otherwise similar i.e., produced in the same year and technology class, by the same firm.

Patent reassignment: In a recent paper, Arora, Belenzon, and Suh (2022) argues that science-based inventions have lower transfer costs and offer greater potential gains from trade. As a result, firms are more likely to trade these high-quality patents in the technology market if they deem it more beneficial than commercializing them internally. Given transaction costs, more valuable patents are also more likely to be traded.

In our sample of publicly traded firms, we similarly observe that science-based patents are more likely to be traded, even within firms, as shown in Column 3 of Table 8, respectively.[16]

[16] Patent reassignment data are obtained from the USPTO Patent Assignment Data set (PAD) (Graham, Marco, & Myers, 2018), which details the transfer of ownership between patent assignees. To ensure that we accurately capture patent assignments relevant to the market for innovation, we exclude employer assignments. Next, we remove assignments associated with administrative events such as corrections and

As with forward citations, the association between science and patent reassignment is virtually unchanged even after controlling for patent value, as Column 4 of Table 8 shows. That is, science-based inventions are more likely to be traded in part because they are more valuable but in part because they have lower transaction costs and have broader applicability than comparable non-science-based inventions.

Patent renewal: We next look at the patent renewal, a commonly used measure of patent value (e.g., Hegde & Sampat, 2009; Schankerman, 1998; Serrano, 2010). Column 5 of Table 8 shows that science-based patents are more likely to be renewed at least once. However, Column 6 shows that once we control for patent value, there is no difference between science-based and non-science-based patents. Because renewal of a patent reflects a firm's assessment of the value of the patent, it implies that science-based inventions are renewed at higher rates, but simply because they are more valuable.

Patent family size: Lastly, we examine another measure of patent value: patent family size, namely the number of patent offices where an invention has been protected, as detailed in Squicciarini et al. (2013). Patent family size serves as a proxy for the private value of a patent to a firm, as companies are likely to bear additional costs for extending protection to other countries only if they view it as beneficial. As shown in Column 7 of Table 8, science-based patents have a 20% larger patent family size compared to other patents. More valuable patents have larger families, but the association between science and family size weakened after we control for patent value, as Column 8 shows.

6 Discussion and Conclusion

Estimating the private economic value of patents is important, yet challenging. By developing a methodology that uses stock market returns to produce a distribution of patent values and not just an estimate of the mean of that distribution, Kogan et al. (2017) has pushed

name changes. We keep conveyance types like "assignment," "govern," and "merger," while excluding those categorized as "correct," "namechg," "missing," "other," "security," and "release."

the field considerably forward, and opened multiple venues for new research.

The KPSS estimates of individual patent values have been used in multiple applications, such as to compare average patent values across different patent types. For example, researchers have used these estimates to gauge whether and when the private returns to produce patents on science-based inventions are higher than patents on non scientific inventions. This literature, however, has produced mixed and specification-dependent evidence.

In this paper, we show that KPSS values should not be used in their current form to compare mean values of different groups of patents, particularly within firms. This is because in producing patent values, KPSS assume that all patents for a given firm-year are drawn from the same value distribution. We show that if the assumption holds, the average values of two groups of patents vary only due to sampling fluctuations. On the other hand, if this assumption is violated, as we show it likely is, then relying on KPSS estimated values introduces a downward-bias by underestimating the value of science-based patents and overestimating the value of non science-based patents. We extend the original KPSS methodology to allow for patents to be drawn from two distinct value distributions, and demonstrate that our approach produces consistent and robust evidence of differences in means between science-based and non-science-based patents.

We find that science-based patents are more valuable. Consistent with this, we also find that they are more likely to be renewed, more likely to be filed in multiple countries, more likely to be traded, and receive more citations from other patents in the future. We conclude that science-based patents are more valuable in part because they represent higher quality inventions. However, even controlling for value, science-based inventions are more useful as inputs into follow-on inventions. They are also likely to be easier to trade. Together, this implies that the social value of science-based inventions may be even higher than the private value. That is, the societal science premium may be even greater than the private one.

References

Ahmadpoor, M., & Jones, B. F. (2017). The dual frontier: Patented inventions and prior scientific advance. *Science*, *357*(6351), 583–587.

Aigner, D., Lovell, C. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of econometrics*, *6*(1), 21–37.

Arora, A., Belenzon, S., & Dionisi, B. (2023). First-mover advantage and the private value of public science. *Research Policy*, *52*(9), 104867.

Arora, A., Belenzon, S., & Sheer, L. (2021a). Knowledge spillovers and corporate investment in scientific research. *American Economic Review*, *111*(3), 871–98.

Arora, A., Belenzon, S., & Sheer, L. (2021b). Matching patents to compustat firms, 1980-2015: Dynamic reassignment, name changes, and ownership structures. *Research Policy*, *50*(5), 104217.

Arora, A., Belenzon, S., & Suh, J. (2022). Science and the market for technology. *Management Science*, *68*(10), 7176–7201.

Arora, A., Cohen, W., Lee, H., & Sebastian, D. (2023). Invention value, inventive capability and the large firm advantage. *Research Policy*, *52*(1), 104650.

Branstetter, L. G., & Ogura, Y. (2005). Is academic science driving a surge in industrial innovation? evidence from patent citations. National Bureau of Economic Research Cambridge, Mass., USA.

Cotropia, C. A., Lemley, M. A., & Sampat, B. (2013). Do applicant patent citations matter? *Research Policy*, *42*(4), 844–854.

Gambardella, A., Harhoff, D., & Verspagen, B. (2008). The value of european patents. *European Management Review*, *5*(2), 69–84.

Graham, S. J., Marco, A. C., & Myers, A. F. (2018). Patent transactions in the marketplace: Lessons from the uspto patent assignment dataset. *Journal of Economics & Management Strategy*, *27*(3), 343–371.

Griliches, Z. (1981). Market value, r&d, and patents. *Economics letters*, *7*(2), 183–187.

Hall, B. H., & Harhoff, D. (2012). Recent research on the economics of patents. *Annual Review of Economics*, *4*(1), 541–565.

Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of Economics*, *36*(1), 16–39.

Harhoff, D., Scherer, F. M., & Vopel, K. (2003). Citations, family size, opposition and the value of patent rights. *Research policy*, *32*(8), 1343–1363.

He, J., Ren, X., & Tian, X. (2023). Do short sellers affect corporate innovation? evidence from a policy experiment. *The Review of Corporate Finance Studies*, cfad007.

Hegde, D., & Sampat, B. (2009). Examiner citations, applicant citations, and the private value of patents. *Economics Letters*, *105*(3), 287–289.

Jondrow, J., Lovell, C. K., Materov, I. S., & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of econometrics*, *19*(2-3), 233–238.

Keum, D. D. (2021). Innovation, short-termism, and the cost of strong corporate governance. *Strategic Management Journal*, *42*(1), 3–29.

Kline, S. J., & Rosenberg, N. (1986). An Overview of Innovation. In N. Rosenberg & R. Landau (Eds.), *The positive sum strategy: Harnessing technology for economic growth* (pp. 266–290). Washington, DC: National Academies Press.

Kogan, L., Papanikolaou, D., Seru, A., & Stoffman, N. (2017). Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics*, *132*(2), 665–712.

Krieger, J. L., Schnitzer, M., & Watzinger, M. (2024). Standing on the shoulders of science. *Strategic Management Journal*, *45*(9), 1670–1695.

Li, Q., Ma, M. S., & Shevlin, T. (2021). The effect of tax avoidance crackdown on corporate innovation. *Journal of Accounting and Economics*, *71*(2-3), 101382.

Marx, M., & Fuegi, A. (2020). Reliance on science: Worldwide front-page patent citations to scientific articles. *Strategic Management Journal*, *41*(9), 1572–1594.

Nagaoka, S., & Walsh, J. P. (2009). *The r&d process in the us and japan: Major findings from the rieti-georgia tech inventor survey*. Research Institute of Economy, Trade and Industry (RIETI).

Narin, F., Hamilton, K. S., & Olivastro, D. (1997). The increasing linkage between us technology and public science. *Research policy*, *26*(3), 317–330.

Nelson, R. (1962). The link between science and invention: The case of the transistor. In *The rate and direction of inventive activity: Economic and social factors* (pp. 549–584). Princeton University Press.

Nicholas, T. (2008). Does innovation cause stock market runups? evidence from the great crash. *American Economic Review*, *98*(4), 1370–1396.

Owen, D. B. (1980). A table of normal integrals: A table. *Communications in Statistics-Simulation and Computation*, *9*(4), 389–419.

Pakes, A., Simpson, M., Judd, K., & Mansfield, E. (1989). Patent renewal data. *Brookings papers on economic activity. Microeconomics*, *1989*, 331–410.

Papadopoulos, A. (2015). The half-normal specification for the two-tier stochastic frontier model. *Journal of Productivity Analysis*, *43*, 225–230.

Papadopoulos, A. (2021). The two-tier stochastic frontier framework (2tsf): Measuring frontiers wherever they may exist. In *Advances in efficiency and productivity analysis* (pp. 163–194). Springer.

Priem, J., Piwowar, H., & Orr, R. (2022). Openalex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv preprint arXiv:2205.01833*.

Roach, M., & Cohen, W. M. (2013). Lens or prism? patent citations as a measure of knowledge flows from public research. *Management Science*, *59*(2), 504–525.

Rosenberg, N. (1990). Why do firms do basic research (with their own money)? *Research Policy*, *19*(2), 165–174.

Sanders, B. S., Rossman, J., & Harris, L. J. (1958). The economic impact of patents. *Patent, Trademark and Copyright Journal of Research and Education*, *2*, 340.

Schankerman, M. (1998). How valuable is patent protection? estimates by technology field. *the RAND Journal of Economics*, 77–107.

Serrano, C. J. (2010). The dynamics of the transfer and renewal of patents. *The RAND Journal of Economics*, *41*(4), 686–708.

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.-J., & Wang, K. (2015). An overview of microsoft academic service (mas) and applications. In *Proceedings of the 24th international conference on world wide web* (pp. 243–246).

Sorenson, O., & Fleming, L. (2004). Science and the diffusion of knowledge. *Research Policy*, *33*(10), 1615–1634.

Squicciarini, M., Dernis, H., & Criscuolo, C. (2013). Measuring patent quality: Indicators of technological and economic value. *OECD Science, Technology and Industry Working Papers*, *2013*(3), 0_1.

Trajtenberg, M. (1990). A penny for your quotes: Patent citations and the value of innovations. *The Rand journal of economics*, 172–187.

Table 2: Descriptive statistics of key patent indicators

	Obs.	Mean	Std. Dev.	10%	50%	90%
Patent Value: KPSS (const. USD mn):	1,448,025	17.516	42.109	1.450	7.435	37.406
Patent Value: Sci&Non-Sci (separate) (const. USD mn)	1,448,025	18.181	43.018	1.507	7.735	39.065
Patent Value: Sci&Non-Sci (mixed) (const. USD mn)	1,448,025	17.232	41.783	1.419	7.230	36.757
Patent forward cites	1,444,640	14.406	50.331	0.000	5.000	28.000
I[Top 1% of cited patents] x 100	1,444,640	1.284	11.257	0.000	0.000	0.000
I[Reassignment]*100	1,448,025	7.382	26.148	0.000	0.000	0.000
Renewd (at least once)	1,266,884	0.844	0.363	0.000	1.000	1.000
Patent Family Size	1,444,640	3.801	4.571	1.000	2.000	8.000
Science Dummy	1,448,025	0.240	0.427	0.000	0.000	1.000

Note: We defined "science-based patents" as those ranking in the top three quartiles for the number of non-patent literature (NPL) citations within a specific IPC class and year, provided they cite at least one NPL. *Patent Value: KPSS (const. USD mn)* refers to the private value of a patent estimated using the KPSS methodology. The *Patent Value: Sci&Non-Sci (separate) (const. USD mn)* is the patent's private value estimated using separate distribution assumptions for science-based and other patents. However, patent value is equally distributed among all patents granted on the science patent days (when at least one science-based patent is granted on a given grant date). Meanwhile, in the *Patent Value: Sci&Non-Sci (mixed) (const. USD mn)*, we differentiate science-based patents from non-science-based patents using our modified formula. Patent forward cites are the total number of forward citations garnered by a patent over five years. *I[Top 1% of cited patents]* is a binary variable equal to 1 if the patent is in the top 1% of cited patents. *I[Reassignment]* is an indicator variable equal to 1 if the patent has ever been reassigned in the USPTO PAD, and zero otherwise. *Renewal (at least once)* is a dummy variable indicating whether the patent maintenance fees were paid at the end of the 4th year. *Patent Family Size* is the number of patent offices at which a given invention is protected. The data is sourced from the OECD patent quality indicators database (Squicciarini, Dernis, & Criscuolo, 2013). *** p<0.01, ** p<0.05, * p<0.1

Table 3: Mean comparisons of key indicators for science-based and non-science patents

	(Science-based Patents)		(Other Patents)		(Difference)
	(1)		(2)		(1-2)
	mean	sd	mean	sd	b
Patent Value: KPSS (const. USD mn):	21.007	48.281	16.413	39.896	4.594***
Patent Value: Sci&Non-Sci (separate) (const. USD mn)	22.671	52.117	16.761	39.603	5.910***
Patent Value: Sci&Non-Sci (mixed) (const. USD mn)	22.563	51.956	15.546	37.848	7.017***
Patent forward cites	24.991	80.007	11.055	35.523	13.936***
I[Top 1% of cited patents] x 100	2.525	15.689	0.890	9.394	1.635***
I[Reassignment]*100	9.140	28.818	6.827	25.220	2.313***
Renewd (at least once)	0.845	0.362	0.844	0.363	0.001
Patent Family Size	5.086	6.237	3.394	3.808	1.692***
Observations	347,901		1,100,124		1,448,025

Note: We defined "science-based patents" as those ranking in the top three quartiles for the number of non-patent literature (NPL) citations within a specific IPC class and year, provided they cite at least one NPL. *Patent Value: KPSS (const. USD mn)* refers to the private value of a patent estimated using the KPSS methodology. The *Patent Value: Sci&Non-Sci (separate) (const. USD mn)* is the patent's private value estimated using separate distribution assumptions for science-based and other patents. However, patent value is equally distributed among all patents granted on the science patent days (when at least one science-based patent is granted on a given grant date). Meanwhile, in the *Patent Value: Sci&Non-Sci (mixed) (const. USD mn)*, we differentiate science-based patents from non-science-based patents using our modified formula. Patent forward cites are the total number of forward citations garnered by a patent over five years. *I[Top 1% of cited patents]* is a binary variable equal to 1 if the patent is in the top 1% of cited patents. *I[Reassignment]* is a binary variable equal to 1 if the patent has ever been reassigned in the USPTO PAD, and zero otherwise. *Renewal (at least once)* is a dummy variable indicating whether the patent maintenance fees were paid at the end of the 4th year. *Patent Family Size* is the number of patent offices at which a given invention is protected. The data is sourced from the OECD patent quality indicators database (Squicciarini, Dernis, & Criscuolo, 2013). *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 4: Patent private value and technical value

	Forward citations				Breakthrough			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(Pat Value: KPSS)		0.0830***				0.1489***		
		(0.0012)				(0.0131)		
Log(Pat Value: Sci&Non-Sci (separate))			0.0880***				0.1693***	
			(0.0012)				(0.0133)	
Log(Pat Value: Sci&Non-Sci (mixed))				0.0957***				0.2026***
				(0.0012)				(0.0131)
Constant	1.8260***	0.5122***	0.4302***	0.3137***	1.2830***	-1.0732***	-1.4018***	-1.9175***
	(0.0009)	(0.0189)	(0.0191)	(0.0188)	(0.0092)	(0.2071)	(0.2098)	(0.2066)
Avg of DV	1.826	1.826	1.826	1.826	1.283	1.283	1.283	1.283
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
4-digit IPC Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.2080	0.2105	0.2108	0.2114	0.0418	0.0419	0.0419	0.0420
N	1,443,253	1,443,253	1,443,253	1,443,253	1,443,253	1,443,253	1,443,253	1,443,253

Note: The dependent variable in columns (1-4) is the *Log (1 + patent Forward citation)*, which is the natural log of the patent's forward citations garnered over a five-year period. The dependent variable in columns (5-8) is a binary variable equal to one if the focal patent received a number of forward citations in the top 99th percentile among the patents granted in the same year and within the same patent class, as sourced from Squicciarini et al. (2013). The explanatory variable in columns (2) and (6) is the natural log transformation of the patent's private value estimated using the KPSS method. The explanatory variable in columns (3) and (7) is the natural log transformation of the private value of patents: *Patent Value: Sci&Non-Sci (separate)* and *Patent Value: Sci&Non-Sci (mixed)*, as shown in columns (4) and (8). For further details, see Table 9. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Patent private value and science premium: KPSS methodology and modified estimation approach

	Pat Value: KPSS Estimates			Pat Value: Sci&Non-Sci (separate)			Pat Value: Sci&Non-Sci (mixed)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Science Dummy	0.1132***	0.0120***	0.0080***	0.1593***	0.0524***	0.0423***	0.2384***	0.1373***	0.1334***
	(0.0025)	(0.0026)	(0.0016)	(0.0025)	(0.0026)	(0.0016)	(0.0025)	(0.0026)	(0.0016)
Constant	15.7919***	15.8162***	15.8182***	15.8180***	15.8438***	15.8473***	15.7374***	15.7617***	15.7637***
	(0.0012)	(0.0011)	(0.0007)	(0.0012)	(0.0012)	(0.0007)	(0.0012)	(0.0011)	(0.0007)
Avg DV	15.819	15.819	15.820	15.856	15.856	15.857	15.795	15.795	15.796
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
4-digit IPC Fixed Effects	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Firm Fixed Effects	No	No	Yes	No	No	Yes	No	No	Yes
R^2	0.066	0.137	0.680	0.068	0.138	0.687	0.071	0.142	0.681
N	1,448,025	1,447,248	1,446,634	1,448,025	1,447,248	1,446,634	1,448,025	1,447,248	1,446,634

Note: The dependent variable in columns (1-3) is the natural log transformation of the patent private value estimated using stock market reactions following the KPSS methodology. The dependent variable in columns (4-6) is the natural log transformation of the patent private value: *Patent Value: Sci&Non-Sci (separate)*, estimated using separate distribution assumptions for science-based and other patents. However, the patent value is equally distributed among all patents granted on the science patent day—days when at least one science-based patent is granted. In columns (7-9), the dependent variable is *Patent Value: Sci&Non-Sci (mixed)*, where we further differentiate science-based patents from non-science-based patents using our modified formula and estimate the patent value for the days when both types of patents were granted. The "science-based patents" are defined as those ranking in the top three quartiles for the number of non-patent literature (NPL) citations within a specific IPC class and year, provided they cite at least one NPL. For further details, see Table 9. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Patent private value and science: Alternative definition of science-based patents

	Defining science via number of NPL citations		
	(1)	(2)	(3)
	Top 3 Quartile	Above Median	Confidence score >8
Science Dummy	0.1334***		
	(0.0016)		
Science Dummy(median)		0.1654***	
		(0.0018)	
Science Dummy (conf9)			0.1608***
			(0.0016)
Constant	15.7637***	15.7656***	15.7571***
	(0.0007)	(0.0007)	(0.0007)
Year Fixed Effects	Yes	Yes	Yes
4-digit IPC Fixed Effects	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes
R^2	0.681	0.681	0.682
N	1,446,634	1,446,634	1,446,634
$\gamma\_science$	0.017	0.017	0.017
γ_Other	0.022	0.024	0.023

Note: The dependent variable is the private value of patent: *Patent Value: Sci&Non-Sci (mixed)*. We defined "science-based patents" as those ranking in the top three quartiles for the number of non-patent literature (NPL) citations within a specific IPC class and year, provided they cite at least one NPL. We further refined our definition of science-based patents to include only those patents that cite NPL above the median number of citations in column (2). In column (3), we include only patents with a publication citation confidence score above 8. For further details, see Table 9. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Patent private value and science: Technological Sector

	Private Value and Science				
	(1)	(2)	(3)	(4)	(5)
	All	Life Science	Chemicals	ICT	Other
Science Dummy	0.1334***	0.0984***	0.1215***	0.1305***	0.1401***
	(0.0016)	(0.0056)	(0.0042)	(0.0020)	(0.0039)
Constant	15.7637***	16.0734***	16.1485***	15.6489***	15.7382***
	(0.0007)	(0.0035)	(0.0021)	(0.0010)	(0.0012)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
4-digit IPC Fixed Effects	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes
R^2	0.681	0.793	0.729	0.686	0.666
N	1,446,634	108,901	171,121	678,422	411,544

Note: The dependent variable is the natural log transformation of the private value of patent: *Patent Value: Sci&Non-Sci (mixed)*, estimated using separate distribution assumptions for science-based and other patents. We defined “science-based patents” as those ranking in the top three quartiles for the number of non-patent literature (NPL) citations within a specific IPC class and year, provided they cite at least one NPL. All other definitions are identical to those in Table 9. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Patent private value, science and other indicators of quality

	Log(1+Forward cit.)		I[Reassignment]*100		Renewal(at least once)		Family Size (Poisson)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Science Dummy	0.3782***	0.3673***	1.2978***	1.1851***	0.0037***	0.0011	0.1822***	0.1748***
	(0.0025)	(0.0025)	(0.0572)	(0.0573)	(0.0007)	(0.0007)	(0.0021)	(0.0021)
Log(Pat Value: Sci&Non-Sci (mixed))		0.0821***		0.8444***		0.0205***		0.0581***
		(0.0012)		(0.0297)		(0.0004)		(0.0011)
Constant	1.7350***	0.4407***	7.0608***	-6.2498***	0.8433***	0.5190***	1.4801***	0.5535***
	(0.0010)	(0.0186)	(0.0244)	(0.4669)	(0.0003)	(0.0062)	(0.0010)	(0.0178)
Avg of DV	1.826	1.826	7.373	7.373	0.844	0.844	3.800	3.800
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
4-digit IPC Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.223	0.225	0.072	0.072	0.214	0.215		
R^2_P							0.239	0.240
N	1,443,253	1,443,253	1,446,634	1,446,634	1,265,742	1,265,742	1,443,253	1,443,253

Note: The dependent variable in columns (1) and (2) is *Log (1 + patent forward citations)*, which represents the natural logarithm of the patent's forward citations garnered over a five-year period. The dependent variable in columns (3) and (4) is *I[Reassignment]*, an indicator variable equal to 1 if the patent has ever been reassigned in the USPTO Patent Assignment Dataset (PAD), and zero otherwise. The dependent variable in columns (5) and (6), *Renewal (at least once)* is a dummy variable indicating whether the patent maintenance fees were paid at the end of the 4th year. In columns (7) and (8), *Patent Family Size* represents the number of patent offices at which a given invention is protected. The data is sourced from the OECD patent quality indicators database (Squicciarini, Dernis, & Criscuolo, 2013). We defined "science-based patents" as those ranking in the top three quartiles for the number of non-patent literature (NPL) citations within a specific IPC class and year, provided they cite at least one NPL. All other definitions are identical to those in Table 9. Robust standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Variable name and description

Indicator Name	Description
Science-based patents	Patents ranked in the top three quartiles for non-patent literature (NPL) citations within a specific IPC class and year, provided they cite at least one NPL.
Science Dummy	A dummy variable equal to 1 if the patent is science-based, and 0 otherwise.
Patent Value: KPSS	The private value of a patent estimated using the KPSS methodology described in Section 2.2.1.
Patent Value: Sci & Non-Sci (separate)	The patent's private value, estimated using separate distribution assumptions for science-based and other patents. However, patent value is distributed equally among all patents granted on days when at least one science-based patent is granted. we estimate the signal-to-noise ratio for each type of patent from Equation 12 and patent value estimates as described in 3.
Patent Value: Sci & Non-Sci (mixed)	The patent's private value, estimated using separate distribution assumptions for science-based and other patents. we use the method described in Section 3.2 to estimate the patent private value on science days when both types of patents are present (at least one science-based and one non-science-based patent).
Patent forward cites	forward citations received by the focal patent within 5 years from its issuing, sourced from the OECD patent quality indicators database (Squicciarini, Dernis, & Criscuolo, 2013).
Breakthrough or Top 1% cited patent	A dummy variable equal to 1 if the patent belongs to the top 1% of highly cited patents within five years after publication in a given year and technological class. The data is sourced from the OECD patent quality indicators database (Squicciarini, Dernis, & Criscuolo, 2013).
Reassignment Dummy	A binary variable equal to 1 if the patent has ever been reassigned. Patent reassignment data obtained from the USPTO Patent Assignment Data set (PAD) (Graham, Marco, & Myers, 2018), which details the transfer of ownership between patent assignees. Employer assignments are excluded to ensure we capture assignments relevant to the innovation market. Administrative events such as corrections and name changes are removed. Conveyance types like "assignment," "govern," and "merger" are kept, while "correct," "namechg," "missing," "other," "security," and "release" are excluded.
Renewal (at least once)	A dummy variable indicating whether the patent maintenance fees were paid at the end of the 4th year.
Patent Family Size	Patent family size is the number of patent offices at which a given invention is protected. The data is sourced from the OECD patent quality indicators database (Squicciarini, Dernis, & Criscuolo, 2013).

Appendix A

Estimating the value of patents with single patent type: the KPSS estimation

A1 KPSS patent value estimation methodology

KPSS make the following assumptions:

1. x_j is normally distributed truncated at zero $\mathcal{N}(0, \sigma^2_{xft})$
2. ϵ_j is normally distributed $\mathcal{N}(0, \sigma^2_{\varepsilon ft})$
3. σ^2_{xft} and $\sigma^2_{\varepsilon ft}$ vary across firms and across time but in constant proportions
4. x_j and ϵ_j are independent.

Under these assumptions, we can calculate $\mathbb{E}[x_j|r_j]$ using the distribution of x_j conditional on r_j. By independence, the joint density of $f(x_j, \epsilon_j)$ is product of their density:

$$f(x_j, \epsilon_j) = \frac{1}{(\pi\sigma_{xft}\sigma_{\varepsilon ft})} \exp\Big[\frac{-1}{2\sigma^2_{xft}} x_j^2 - \frac{1}{\sigma^2_{\varepsilon ft}} \epsilon_j^2\Big]; x_j > 0 \tag{15}$$

using the transformation $\epsilon_j = r_j - x_j$:

$$f(x_j, r_j) = \frac{1}{(\pi\sigma_{xft}\sigma_{\varepsilon ft})} \exp\Big[\frac{-1}{2\sigma^2_{vft}} x_j^2 - \frac{1}{2\sigma^2_{\varepsilon ft}} (r_j - x_j)^2\Big] \tag{16}$$

Using Aigner, Lovell, and Schmidt (1977), the density function of $f(r_j)$ is given by:

$$f(r_j) = \frac{2}{(\sqrt{2\pi}\sigma)} (\Phi(\frac{r_j\lambda}{\sigma})) \exp\Big[\frac{-1}{2\sigma^2}(r_j)^2\Big] \tag{17}$$

where $\sigma = \sigma^2_{xft} + \sigma^2_{\varepsilon ft}$, $\lambda = \frac{\sigma_{xft}}{\sigma_{\varepsilon ft}}$, and $\Phi(.)$ is CDF of standard normal distribution. As shown by Jondrow, Lovell, Materov, and Schmidt (1982), letting $\sigma^2_* = \frac{\sigma^2_{xft}\sigma^2_{\varepsilon ft}}{\sigma^2}$ we can calculate the conditional distribution and then expected values from:

$$f(x_j|r_j) = \frac{1}{(\Phi(\frac{r_j\lambda}{\sigma}))} \frac{1}{(\sqrt{2\pi}\sigma_*)} \exp\Big[\frac{-1}{2\sigma^2_*}(-x_j + \frac{\sigma^2_{xft} r_j}{\sigma^2})^2\Big] \tag{18}$$

The distribution of $f(x_j|r_j)$ is the same as a Normal distribution with $\mathcal{N}(\mu_*, \sigma^2_*)$ multiplied by $\frac{1}{(\Phi(\frac{r_j\lambda}{\sigma}))}$, where $\mu_* = \frac{\sigma^2_{xft} r_j}{\sigma^2_{\varepsilon ft} + \sigma^2_{xft}}$. The density is equivalent to the $\mathcal{N}(\mu_*, \sigma^2_*)$ distribution truncated at zero.

Using the expectation formula for truncated normal distribution:

$$\mathbb{E}[x_j|r_j] = \mu_* + \sigma_* \frac{\phi(R_j)}{1 - \Phi(R_j)} \tag{19}$$

where $R_j = \mu_*/\sigma^2_* = -\sqrt{\delta_j}\frac{r_j}{\sigma_{\varepsilon ft}}$ and $\delta_j = \frac{\sigma^2_{xft}}{\sigma^2_{\varepsilon ft} + \sigma^2_{xft}}$. Thus, μ_* can be written as $\delta_j r_j$ and σ^2_* becomes $\sqrt{\delta_j}\sigma_{\varepsilon ft}$, leading to the following formula:

$$\mathbb{E}[x_j|r_j] = \delta_j r_j + \sqrt{\delta_j}\sigma_{\varepsilon ft}\frac{\phi(R_j)}{1-\Phi(R_j)} \tag{20}$$

To estimate the value of patents from the formula, the parameters δ_j and $\sigma_{\varepsilon ft}$ need to be estimated. KPSS assume that δ_j, which is the ratio of the variance of x_j to the sum of the variance of x_j and ϵ_j, is constant across firms and over time. To compute δ, they estimate:

$$ln(r^l_{fd})^2 = a_0 + a_{ft} + b_d + \gamma I_{fd} + \mu_{fd} \tag{21}$$

where r^l_{fd} is the idiosyncratic return of firm f centered on day d with a window of length l, a_{ft} is the firm-year fixed effect, and b_d is the day-of-week fixed effect. γ is equal to :

$$\mathbb{E}[ln(x_j + \epsilon_j)^2] - \mathbb{E}[ln(\epsilon_j^2)] = \gamma \tag{22}$$

Approximating the distribution of $x_j + \epsilon_j$ as a normal distribution, the square of a standard normal variable is distributed as $\chi^2(1)$.

$$\mathbb{E}\Big[\Big(ln(x_j+\epsilon_j)^2\Big)\Big(\frac{\sigma^2_{\varepsilon ft}+\sigma^2_{xft}}{\sigma^2_{\varepsilon ft}+\sigma^2_{xft}}\Big)\Big] - \mathbb{E}\Big[ln\Big(\epsilon_j^2\Big)\Big(\frac{\sigma^2_{\varepsilon ft}}{\sigma^2_{\varepsilon ft}}\Big)\Big] = \gamma \tag{23}$$

Solving this and adjusting for the truncated variance of x_j leads to:[17]

$$\ln\Big[\frac{\sigma^2_{\varepsilon ft}+\sigma^2_{xft}(1-(\frac{\phi(0)}{1-\Phi(0)})^2)}{\sigma^2_{\varepsilon ft}}\Big] = \gamma \tag{24}$$

$$\Big[\frac{\sigma^2_{\varepsilon ft}+\sigma^2_{xft}(1-(\frac{\phi(0)}{1-\Phi(0)})^2)}{\sigma^2_{\varepsilon ft}}\Big] = e^\gamma \tag{25}$$

simplifying this using $\delta_j = \frac{\sigma^2_{xft}}{\sigma^2_{\varepsilon ft}+\sigma^2_{xft}}$ leads to:

$$\hat{\delta} = 1 - \Bigg(1 + \frac{1}{(1-(\frac{\phi(0)}{1-\Phi(0)})^2)}(e^\gamma - 1)\Bigg)^{-1} = 1 - e^{-\gamma}$$

Next, KPSS need to recover the $\sigma^2_{\varepsilon ft}$. This is done non-parametrically using the sum of squared market-adjusted returns σ_{ft} if m_{ft} is the fraction of trading days with a patent grant in a firm-year.

$$\sigma^2_{\epsilon_{ft}} + m_{ft}\sigma^2_{xft}(1+l) = \sigma^2_{ft}(1+l)$$

[17]The variance of $\mathcal{N}(0,\sigma^2)$ truncated at d is $\sigma^2\lambda(d)(1-\lambda(d))$ where λ is the inverse mill ratio $(\frac{\phi(d)}{(1-\Phi(d)})$. The square of standard Normal is distributed as $\chi^2(1)$.

using further simplification of equation 25 we get:

$$\sigma^2_{\epsilon_{ft}} = \frac{\sigma^2_{ft}(1+l)}{(1+m_{ft}(1+l)(e^{\hat{\gamma}}-1))}$$

A1.1 Distribution of KPSS values

Note that the KPSS values are distributed differently than the distribution of x itself. For a given firm-year, the covariance between $\mathbb{E}[x|r]$ & x is given by $\delta\sigma_x^2 + \sqrt{\delta}\sigma_\varepsilon Cov(\lambda(R),\ x)$. The term $\sqrt{\delta}\sigma_\varepsilon Cov(\lambda(R),\ x) \leq 0$ because $\lambda(R)$ is a decreasing function of r whereas x is an increasing function. Therefore, $Cov(\mathbb{E}[x|r], x) \leq \delta\sigma_x^2$.

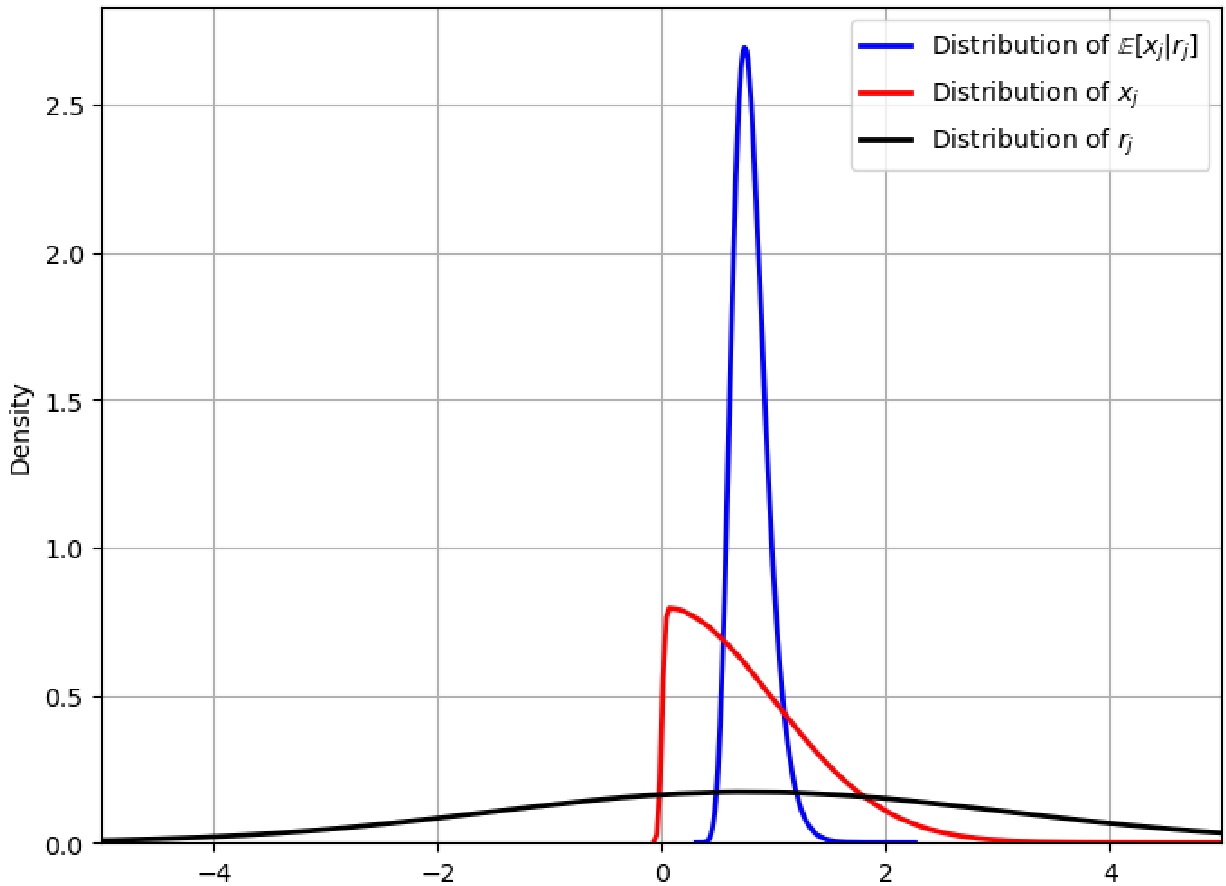

Figure 1: Simulation distribution of x, r, and $\mathbb{E}(x|r)$

Numerical simulations indicate that $\mathbb{E}[x|r]$ has a has a less skewed distribution and smaller variance than x.[18] The median of the distribution of KPSS values is greater than the median of x but also has less probability mass in the tails.

[18]Where $x \sim \mathcal{N}^+(0, \sigma_x^2)$ and $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$, with $\sigma_x^2 = 1$ and $\sigma_\epsilon^2 = 5$. More details on the simulation are provided in Appendix D, where we extend to simulation to two types of patents.

A2 Revisiting the KPSS method to handle multiple patents granted in a single day

$$r = x_1 + x_2 + \varepsilon$$

Where x_1 and x_2 are two patents drawn from the same distribution (e.g., both are science-based or non-science-based).[19]

First, assuming only two patents are granted on a given day. Following the KPSS method as described in Appendix A section A1.

$$ln(r^l_{fd})^2 = a_0 + a_{ft} + b_d + \gamma_1 I_{1fd} + \gamma_2 I_{2fd} + \mu_{fd} \tag{26}$$

I_{1fd} indicates that there is only 1 patent on grant day; similarly, I_{2fd} indicates two patents on the grant day.

$$\ln\Big[\frac{\sigma^2_{\varepsilon ft} + \sigma^2_{xft}}{\sigma^2_{\varepsilon ft}}\Big] = \gamma_1, \ln\Big[\frac{\sigma^2_{\varepsilon ft} + 2 * \sigma^2_{xft}}{\sigma^2_{\varepsilon ft}}\Big] = \gamma_2 \tag{27}$$

This implies

$$1 + \theta^2 = e^{\gamma_1}$$
$$1 + 2\theta^2 = e^{\gamma_2}$$

where $\theta = \frac{\sigma^2_{xft}}{\sigma^2_{\varepsilon ft}}$

We solve for θ^2 from the first equation:

$$\theta^2 = e^{\gamma_1} - 1$$

Substitute θ^2 into the second equation:

$$1 + 2(e^{\gamma_1} - 1) = e^{\gamma_2}$$

Simplify the left-hand side:

$$1 + 2e^{\gamma_1} - 2 = e^{\gamma_2} 2e^{\gamma_1} - 1 = e^{\gamma_2}$$

Thus, we have the relationship between γ_1 and γ_2:

$$e^{\gamma_2} = 2e^{\gamma_1} - 1$$

[19] We dropped the index for patent day etc. to keep the notation simple.

Alternatively, we can express γ_2 in terms of γ_1:

$$\gamma_2 = \ln(2e^{\gamma_1} - 1)$$

Approximation for Small γ

Consider the relationship:

$$e^{\gamma_2} = 2e^{\gamma_1} - 1$$

Using the Taylor expansion for e

$$1 + \gamma_2 = 2(1 + \gamma_1) - 1$$

So, for small γ_1, we can approximate:

For a more exact representation, the original expression $\gamma_2 = \ln(2e^{\gamma_1} - 1)$ is the most accurate, but the approximation $\gamma_2 \approx 2\gamma_1$ provides a useful simplification under certain conditions.

We can impose the restriction $\gamma_2 = \ln(2e^{\gamma_1} - 1)$ in equation 26 which becomes

$$ln(r^l_{fd})^2 = a_0 + a_{ft} + b_d + \gamma_1 I_{1fd} + (\ln(2e^{\gamma_1} - 1))I_{2fd} + \mu_{fd} \quad (28)$$

This can be estimated by GMM or non-linear least squares. If we impose the approximation $\gamma_2 = 2\gamma_1$, equation 26 simplifies to

$$ln(r^l_{fd})^2 = a_0 + a_{ft} + b_d + \gamma_1(I_{1fd} + \frac{2}{1}I_{2fd}) + \mu_{fd} \quad (29)$$

This can also extend to n-patent granted on a given day.

$$ln(r^l_{fd})^2 = a_0 + a_{ft} + b_d + \gamma n + \mu_{fd} \quad (30)$$

A2.1 Correcting the bias in KPSS estimates if more than one patent is granted on a single day

Suppose n patents are granted on a particular day, where n is a random variable. As before, we assume that the probability that any patent is granted in a given day is $p > 0$. We show that the KPSS version of equation 29 yields an upward biased estimate, where the bias is given by $\gamma_1(1 - N)$, where $N = E(n|I_{1fd} = 1) = np$. Put differently, one needs to divide the KPSS estimate of γ_1 by the expected number of patents conditional on a patent being granted.

To see this, consider a simplified version of of equation 29, where we dispense with the subscripts to reduce notational clutter.

$$Y = a_0 + \gamma n + \zeta \tag{31}$$

where Y is the square of the log returns, and n is the number patents granted, $n = 0, 1, 2, ...$ Consider the analog of the KPSS estimating equation

$$Y = a_0 + \beta I + \zeta \tag{32}$$

where I is a dummy variable that takes value one if there is a patent granted and zero otherwise.

Then the expected value of b, the OLS estimate of β is given by

$$\begin{aligned} \mathbb{E}(b) &= \frac{Cov(Y, I)}{Var(I)} \\ Cov(Y, I) &= E(YI) - E(Y)E(I) = \gamma np - (a_0 + \gamma np)p \\ \implies \mathbb{E}(b) &= \gamma N \end{aligned} \tag{33}$$

This implies that KPSS estimates of γ should be divided by the mean number of patents per day, conditional on patents being granted.

Figure 2: Distribution of average patent value by number of patents on the grant day

Notes: The following plot shows the average patent value calculated using the KPSS assumption (where the total invention value of all patents granted on a given day follows a half-normal distribution) compared to a scenario where we assume a patent value distribution and adjust for the number of patents when calculating the signal-to-noise ratio. The x-axis represents the number of patents granted on a given day. The KPSS method tends to overestimate the patent value compared to the average when the number of patents granted is low, while the modified approximation, which adjusts for the number of patents, underestimates the patent value compared to the average in the same scenario.

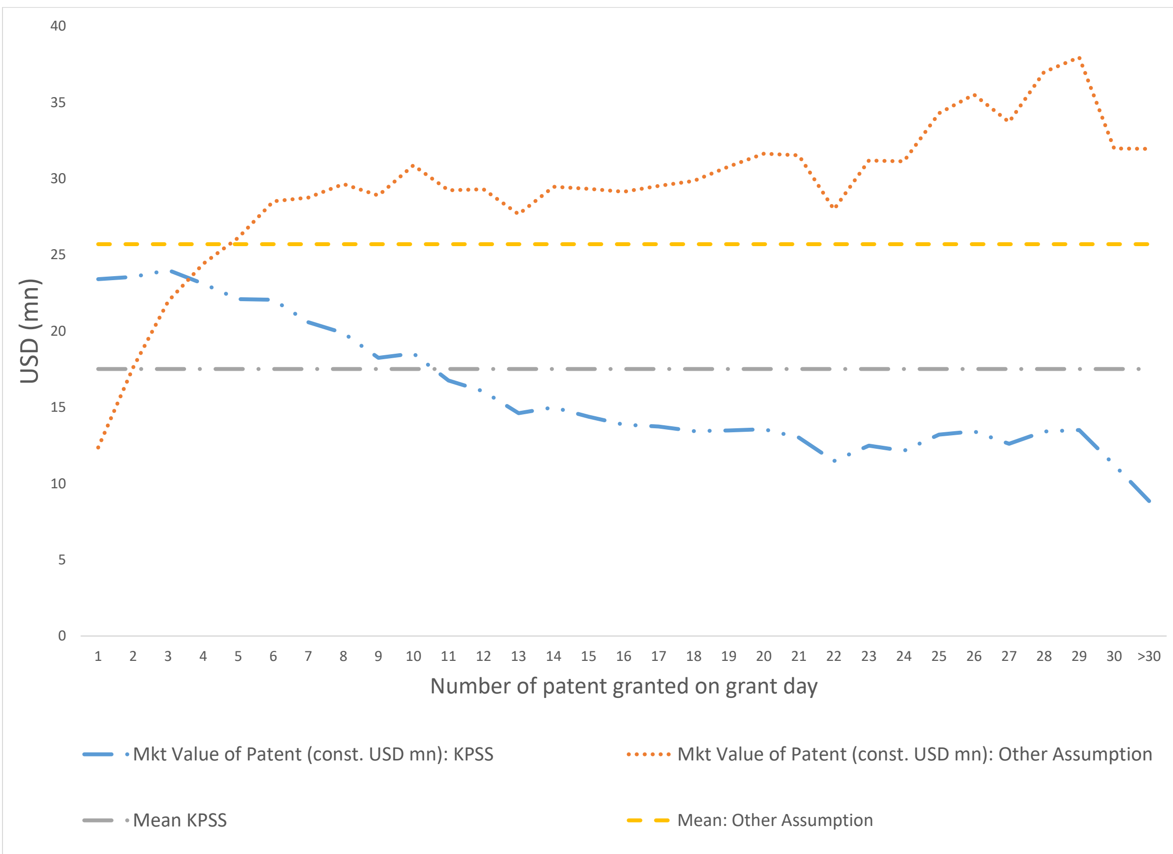

Table A1: Patent private value estimation and KPSS assumptions

	Forward citations						Breakthrough					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Log(Pat Value:KPSS)		0.0534***			0.0830***			0.0586***			0.1489***	
		(0.0008)			(0.0012)			(0.0084)			(0.0131)	
Log(Patent value: New Assump.)			0.0109***			0.1030***			-0.0795***			0.1661***
			(0.0007)			(0.0015)			(0.0081)			(0.0162)
Constant	1.8260***	0.9812***	1.6482***	1.8260***	0.5122***	0.1542***	1.2833***	0.3567***	2.5737***	1.2830***	-1.0732***	-1.4132***
	(0.0009)	(0.0127)	(0.0121)	(0.0009)	(0.0189)	(0.0240)	(0.0093)	(0.1333)	(0.1332)	(0.0092)	(0.2071)	(0.2622)
Avg of DV	1.826	1.826	1.826	1.826	1.826	1.826	1.283	1.283	1.283	1.283	1.283	1.283
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
4-digit IPC Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
R^2	0.1317	0.1345	0.1318	0.2080	0.2105	0.2105	0.0052	0.0052	0.0053	0.0418	0.0419	0.0419
N	1,443,868	1,443,868	1,443,868	1,443,253	1,443,253	1,443,253	1,443,868	1,443,868	1,443,868	1,443,253	1,443,253	1,443,253

Note: The dependent variable in odd columns (1-6) is the *Log (1 + patent Forward citation)*, which is the natural log of the patent's forward citations garnered over a five-year period. The dependent variable in even columns (7-12) is a binary variable equal to one if the focal patent received a number of forward citations in the top 99th percentile among the patents granted in the same year and within the same patent class, as sourced from Squicciarini et al. (2013). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix B
Estimating the value of patents with multiple patent types

B1 Estimating the Value of Patents with Multiple Patent Types

We make a similar assumption to KPSS, with a slight variation:

- x_{sfti} & x_{nftk} denote science-based and non science patent values respectively for firm f and time t.
- x_{sfti} & x_{nftk} are normally distributed truncated at zero but have different variance ($\mathcal{N}^+(0, \sigma^2_{xsft})$ and $\mathcal{N}^+(0, \sigma^2_{xnft})$).
- ε_j is normally distributed $\mathcal{N}(0, \sigma^2_{\varepsilon ft})$.
- σ^2_{xft}, σ^2_{xnft} and $\sigma^2_{\varepsilon ft}$ vary across firms and across time but in constant proportions.
- $\delta_{sft} = \frac{\sigma^2_{xsft}}{\sigma^2_{\varepsilon ft}+\sigma^2_{xsft}} = \delta_s$ and $\delta_{nft} = \frac{\sigma^2_{xnft}}{\sigma^2_{\varepsilon ft}+\sigma^2_{x}nft} = \delta_n$
- x_{sift}, x_{njft} and ϵ_{ft} are independent for a given firm and time.

To derive the $\mathbb{E}[x_j|r_j]$ it is useful to analyze the sum of random variables. Let $r = u+v+w$ where v and w are normally distributed, truncated at zero, but have different variances: $v \sim \mathcal{N}^+(0, \sigma_1^2)$ and $w \sim \mathcal{N}^+(0, \sigma_2^2)$ and u is normally distributed $u \sim \mathcal{N}(0, \sigma_u^2)$).[20] We follow Papadopoulos (2015), which entails these steps:

- Derive the distribution of $z = v + w$

$$f_Z(z) = \frac{4}{s_h}\phi(z/s_h)\left[\Phi\left(\frac{\sigma_1}{\sigma_2}(z/s_h)\right) + \Phi\left(\frac{\sigma_2}{\sigma_1}(z/s_h)\right) - 1\right] \tag{34}$$

- Derive the distribution of r

$$f_r(r) = \frac{2}{s}\phi(r/s)\left[\Phi\left(\frac{ar}{\sqrt{1+\lambda_1^2}}\right) - 2T\left(\frac{ar}{\sqrt{1+\lambda_1^2}}, \lambda_1\right) + \Phi\left(\frac{ar}{\sqrt{1+\lambda_2^2}}\right) - 2T\left(\frac{ar}{\sqrt{1+\lambda_2^2}}, \lambda_2\right)\right] \tag{35}$$

  Where:

  - $s^2 = \sigma_u^2 + \sigma_1^2 + \sigma_2^2$
  - $s_h^2 = \sigma_2^2 + \sigma_1^2$

[20] The relationship to patent values is readily apparent when r represents market returns, u represents the noise term, u represents value of science-based patents, and v represents the value of non science-based patents. This substitution simplifies notation, particularly subscripts.

- $\lambda_1 = \frac{s\lambda}{\sigma_u}$ & $\lambda_2 = \frac{s}{\lambda\sigma_u}$ where $\lambda = \frac{\sigma_1}{\sigma_2}$
- $a = \frac{s_h}{s\sigma_u}$

- Derive the conditional density $f(v|r)$

$$f_{v|r}(v|r) = A^{-1}\frac{2}{\omega_v}\phi\left(\frac{v}{\omega_v} - \frac{r}{\omega_1}\right)\Phi\left(\lambda_1\frac{r-v}{\omega_1}\right) \tag{36}$$

Where:

- $\frac{\sigma_1}{ss_2} = \frac{\theta_1}{s\sqrt{1+\theta_2^2}} = \frac{1}{\omega_1}$
- $\frac{\theta_2}{s_2} = \frac{\theta_2}{\theta_1}\sqrt{1+\theta_1^2+\theta_2^2}\frac{1}{\omega_1} = \frac{\lambda_1}{\omega_1}$
- $\omega_v = \frac{\sigma_1 s_2}{s}$

- Derive the condition expectation $\mathbb{E}(v|r)$

$$E(v|r) = \frac{\theta_1^2}{(1+\theta_1^2+\theta_2^2)}r + \frac{\left\{2\frac{(1+\theta_2^2)\sigma_u^2}{\omega_1}\phi(r/\omega_1)\Phi\left(\frac{\lambda_1}{\omega_1}r\right) - \frac{2\theta_1^2\sigma_u^2}{\omega_2}\phi\left(\frac{r}{\omega_2}\right)\Phi\left(\frac{\lambda_2}{\omega_2}r\right)\right\}}{\left[\Phi\left(\frac{r}{\omega_1}\right) - 2T\left(\frac{r}{\omega_1},\lambda_1\right) + \Phi\left(\frac{r}{\omega_2}\right) - 2T\left(\frac{r}{\omega_2},\lambda_2\right)\right]} \tag{37}$$

We will use the following parameterization, where:

- $\theta_1 = \frac{\sigma_1}{\sigma_u}$
- $\theta_2 = \frac{\sigma_2}{\sigma_u}$
- $s = \sqrt{\sigma_u^2+\sigma_1^2+\sigma_2^2} = \sigma_u\sqrt{1+\theta_1^2+\theta_2^2}$, with
- $\omega_1 = \frac{s\sqrt{1+\theta_2^2}}{\theta_1}$
- $\omega_2 = \frac{s\sqrt{1+\theta_1^2}}{\theta_2}$
- $\lambda_1 = \frac{\theta_2}{\theta_1}\sqrt{1+\theta_1^2+\theta_2^2}$
- $\lambda_2 = \frac{\theta_1}{\theta_2}\sqrt{1+\theta_1^2+\theta_2^2}$

- Derive the two signal-to-noise ratios. We can estimate the γ_s and γ_n and recover δ_s as $1-e^{-\gamma_s}$ and δ_n as $1-e^{-\gamma_n}$ from the following regression:

$$ln(r_{fd})^2 = a_0 + a_{ft} + b_d + \gamma_s I_{sfd} + \gamma_n I_{nfd} + \nu_{fd} \tag{38}$$

- Derive the variance of ϵ_j. We can estimate it as follows:

$$\sigma_{\varepsilon ft}^2 + \mu_s\sigma_{sft}^2(1+l) + \mu^N\sigma_{nft}^2(1+l) = \sigma_{ft}^2(1+l) \tag{39}$$

Derive the distribution of $v + w$

$z = v + w$. Both v and w are truncated normal at zero with variance σ_1 & σ_2, so z is the sum of two truncated normal distributions. Further, both v and w are iid so the distribution of z is:

$$F_z(z) = \mathbb{P}(v + w \leq z)$$

$$f_Z(z) = \int_0^z f(z-w) f(w)\, dw$$

$$f_Z(z) = \frac{2}{\pi\sigma_1\sigma_2} \int_0^z \left(\exp\left(-\frac{(z-w)^2}{2\sigma_1^2}\right)\right)\left(\exp\left(-\frac{w^2}{2\sigma_2^2}\right)\right) dw$$

The inside term can written as:

$$\frac{2}{\pi\sigma_1\sigma_2} \exp\left[-\frac{1}{2}\left(\frac{z}{\sigma_1}\right)^2\right] \exp\left[-\frac{1}{2}\frac{s_h^2}{\sigma_1^2\sigma_2^2} w^2 + \frac{z}{\sigma_1^2} w\right]$$

where $s_h^2 = \sigma_1^2 + \sigma_2^2$

Imposing restriction on the domain of w, with the integral between 0 and z, we have to calculate the integral from

$$\int_0^z \left(\exp\left(-\frac{1}{4\delta}\omega^2 + \gamma\omega\right)\right) d\omega$$

Where $\delta = \frac{\sigma_1^2\sigma_2^2}{2s_h^2}$ and $\gamma = \frac{z}{\sigma_1}$

We can write the above expression in the following form:

$$f_Z(z) = \frac{2}{\pi\sigma_1\sigma_2} \exp\left[\frac{1}{2}\left(\frac{z}{s_h^2}\right)^2\right] \sqrt{2\pi\left(\frac{\sigma_1\sigma_2}{s_h}\right)^2} \int_0^z \frac{1}{\sqrt{2\pi\left(\frac{\sigma_1\sigma_2}{s_h}\right)^2}} \exp\left[-\frac{1}{2}\frac{(-(w - \frac{z\sigma_1^2}{s_h^2})^2}{(\frac{\sigma_1\sigma_2}{s_h})^2}\right] dw$$

The density function can be simplified to

$$f_Z(z) = \frac{4}{s_h}\phi(z/s_h)\left[\Phi\left(\frac{\sigma_1}{\sigma_2}(z/s_h)\right) + \Phi\left(\frac{\sigma_2}{\sigma_1}(z/s_h)\right) - 1\right]$$

$$f_Z(z) = \frac{4}{s_h}\phi(z/s_h)\left[\Phi\left(\frac{\sigma_1}{\sigma_2}(z/s_h)\right) - \Phi\left(-\frac{\sigma_2}{\sigma_1}(z/s_h)\right)\right]$$

Derive the distribution of $r = u + v + w$

$r = u + z$ (where u is normal distribution with variance σ_u). The domain of u is $(-\infty, \infty)$, while the domain of z is $(0, \infty)$:

$$F_r(r) = \int_0^\infty \int_{-\infty}^{r-z} f_{u,z}(u,z)dudz$$

$$f_r(r) = \frac{d}{dr}F_r(r) = \int_0^\infty f_{u,z}(r-z,z)dz$$

Since the variables are independent, joint density is the product of the two marginal densities (one normal and one derived in the previous section). We use $1 - \Phi(x) = \Phi(-x)$

$$f_r(r) = \int_0^\infty \frac{1}{\sigma_u}\phi\left(\frac{r-z}{\sigma_u}\right)\frac{4}{s_h}\phi(z/s_h)\left[1 - \Phi\left(-\frac{\sigma_1}{\sigma_2}(z/s_h)\right) - \Phi\left(-\frac{\sigma_2}{\sigma_1}(z/s_h)\right)\right]dz$$

Following Papadopoulos (2015) we get

$$f_r(r) = \frac{2}{s}\phi(r/s)\left[\Phi\left(\frac{ar}{\sqrt{1+\lambda_1^2}}\right) - 2T\left(\frac{ar}{\sqrt{1+\lambda_1^2}}, \lambda_1\right) + \Phi\left(\frac{ar}{\sqrt{1+\lambda_2^2}}\right) - 2T\left(\frac{ar}{\sqrt{1+\lambda_2^2}}, \lambda_2\right)\right]$$

Where:

- $s^2 = \sigma_u^2 + \sigma_1^2 + \sigma_2^2$
- $s_h^2 = \sigma_2^2 + \sigma_1^2$
- $\lambda_1 = \frac{s\lambda}{\sigma_u}$ & $\lambda_2 = \frac{s}{\lambda\sigma_u}$
- $\lambda = \frac{\sigma_1}{\sigma_2}$
- $a = \frac{s_h}{s\sigma_u}$
- T = Owen's T function (Owen, 1980)

Derive the conditional density $f(v|r)$

$$f_{v|r}(v|r) = \frac{f_{v,u+w}(v, u+w)}{F_r(r)}$$

Note: v and w are independent of each other.

$R = \zeta + v$, where $\zeta = u + w$

$$f_{v|r}(v|r) = \frac{f_v(v) f_\zeta(r-v)}{F_r(r)}$$

$$f_{v|r}(v|r) = \frac{\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_1}\exp\left(\frac{-1}{2}(\frac{v}{\sigma_1})^2\right)\frac{2}{s_2}\phi((r-v)/s_2)\Phi\left(\theta_2\frac{r-v}{s_2}\right)}{\frac{2}{s}\phi(r/s)\left[\Phi\left(\frac{ar}{\sqrt{1+\lambda_1^2}}\right) - 2T\left(\frac{ar}{\sqrt{1+\lambda_1^2}},\lambda_1\right) + \Phi\left(\frac{ar}{\sqrt{1+\lambda_2^2}}\right) - 2T\left(\frac{ar}{\sqrt{1+\lambda_2^2}},\lambda_2\right)\right]}$$

Where $s_2^2 = \sigma_2^2 + \sigma_u^2$ and $\theta_2 = \frac{\sigma_2}{\sigma_u}$ & $\theta_1 = \frac{\sigma_1}{\sigma_u}$

$$A = \left[\Phi\left(\frac{ar}{\sqrt{1+\lambda_1^2}}\right) - 2T\left(\frac{ar}{\sqrt{1+\lambda_1^2}},\lambda_1\right) + \Phi\left(\frac{ar}{\sqrt{1+\lambda_2^2}}\right) - 2T\left(\frac{ar}{\sqrt{1+\lambda_2^2}},\lambda_2\right)\right]$$

We can write this in the following form:

$$f_{v|r}(v|r) = A^{-1}\sqrt{\frac{2}{\pi}}\frac{s}{\sigma_1 s_2}\exp\left\{-\frac{1}{2}\left(\frac{v}{\sigma_1}\right)^2 - \frac{1}{2}\left(\frac{r-v}{s_2}\right)^2 + \frac{1}{2}\left(\frac{r}{s}\right)^2\right\}\Phi\left(\theta_2\frac{r-v}{s_2}\right)$$

Following Papadopoulos (2015), we can write
$\frac{\sigma_1}{ss_2} = \frac{\theta_1}{s\sqrt{1+\theta_2^2}} = \frac{1}{\omega_1}$

$\frac{\theta_2}{s_2} = \frac{\theta_2}{\theta_1}\sqrt{1+\theta_1^2+\theta_2^2}\frac{1}{\omega_1} = \frac{\lambda_1}{\omega_1}$

$\omega_v = \frac{\sigma_1 s_2}{s}$

$$f_{v|r}(v|r) = A^{-1}\frac{2}{\omega_v}\phi\left(\frac{v}{\omega_v} - \frac{r}{\omega_1}\right)\Phi\left(\lambda_1\frac{r-v}{\omega_1}\right)$$

The conditional expected value $\mathbb{E}(v|r)$

$$\mathbb{E}(v|r) = \int_0^\infty v f_{v|r}(v|r)dv = \int_0^\infty vA^{-1}\frac{2}{\omega_v}\phi\left(\frac{v}{\omega_v} - \frac{r}{\omega_1}\right)\Phi\left(\lambda_1\frac{r-v}{\omega_1}\right)dv$$

Using the substitution:$v^* = \left(\frac{v}{\omega_v} - \frac{r}{\omega_1}\right)$

$v = \omega_v V^* + (\omega_v/\omega_1)r$

$dv = \omega_v dv^*$

$v = 0 \rightarrow -\frac{r}{\omega_1}$

Substituting:

$$\mathbb{E}(v|r) = (\omega_v/\omega_1)r + (2\omega_v A^{-1}) \int_{-\frac{r}{\omega_1}}^{\infty} (v^*\phi(v^*)\Phi \left\{ - \left(\frac{\lambda_1 r(\omega_v - \omega_1)}{\omega_1^2} + \frac{\lambda_1 \omega_v}{\omega_1} v^* \right) \right\} dv^*$$

Similar to Papadopoulos (2015), this integral can be written as

$$\int_{-\frac{r}{\omega_1}}^{\infty} x\phi(x)\Phi(a+bx)dx$$

From Owen (1980):

$$\int_{c}^{\infty} x\phi(x)\Phi(a+bx) = \frac{b}{\sqrt{1+b^2}}\phi\left(\frac{a}{\sqrt{1+b^2)}}\right)\Phi\left(\frac{c+b(a+bc)}{-\sqrt{1+b^2}}\right) + \phi(c)\Phi(a+bc)$$

Where:

- $a = \frac{-\lambda_1(\omega_v - \omega_1)}{\omega_1^2}$
  $b = \frac{-\lambda_1 \omega_v}{\omega_1}$
  $c = -r/\omega_1$
- $a + bc = \frac{\lambda_1}{\omega_1} r$
- $c + b(a+bc) = -\frac{1}{\omega_1}(1+\omega_2^2)r = \frac{-\theta_1\sqrt{(1+\theta_2^2)}}{s} r$
- $\sqrt{1+b^2} = \frac{s_2}{s}\sqrt{1+\theta_1^2}$

Solving for $\frac{c+b(a+bc)}{\sqrt{1+b^2}} = \frac{\lambda_2}{\omega_2} r$:

- $\frac{a}{\sqrt{1+b^2}} = -\frac{r}{\omega_2}$
- $\frac{b}{\sqrt{1+b^2}} = -\frac{s^2\omega_v}{s_2^2\omega_2}$

Finally, we obtain.

$$E(v|r) = (\omega_v/\omega_1)r + A^{-1}\left\{2\omega_v \frac{\omega_1}{\omega_1}\phi(r/\omega_1)\Phi\left(\frac{\lambda_1}{\omega_1}r\right) - \frac{2s^2\omega_v^2}{s_2^2\omega_2}\phi\left(\frac{r}{\omega_2}\right)\Phi\left(\frac{\lambda_2}{\omega_2}r\right)\right\}$$

Note, this can be further simplified to
$\omega_v/\omega_1 = \frac{\sigma_1^2}{s^2}$

$\omega_v * \omega_1 = \sigma_u^2 + \sigma_2^2$
$\frac{s^2\omega_v^2}{s_2^2} = \sigma_1^2$

$$E(v|r) = \frac{\sigma_1^2}{s^2}r + A^{-1}\left\{(s^2 - \sigma_1^2)g_1 - \sigma_1^2 g_2\right\}$$

using $\frac{a}{\sqrt{1+\lambda_1^2}} = \frac{1}{\omega_1}$ & $\frac{a}{\sqrt{1+\lambda_2^2}} = \frac{1}{\omega_2}$

$$E(v|r) = \frac{\theta_1^2}{(1+\theta_1^2+\theta_2^2)}r + \frac{\left\{2\frac{(1+\theta_2^2)\sigma_u^2}{\omega_1}\phi(R/\omega_1)\Phi\left(\frac{\lambda_1}{\omega_1}r\right) - \frac{2\theta_1^2\sigma_u^2}{\omega_2}\phi\left(\frac{r}{\omega_2}\right)\Phi\left(\frac{\lambda_2}{\omega_2}r\right)\right\}}{\left[\Phi\left(\frac{r}{\omega_1}\right) - 2T\left(\frac{r}{\omega_1},\lambda_1\right) + \Phi\left(\frac{r}{\omega_2}\right) - 2T\left(\frac{r}{\omega_2},\lambda_2\right)\right]} \tag{40}$$

Let
$\theta_1 = \frac{\sigma_1}{\sigma_u}$
$\theta_2 = \frac{\sigma_2}{\sigma_u}$
$s = \sqrt{\sigma_u^2 + \sigma_1^2 + \sigma_2^2} = \sigma_u\sqrt{1+\theta_1^2+\theta_2^2}$

Following Papadopoulos (2015) we get:

- $\omega_1 = \frac{s\sqrt{1+\theta_2^2}}{\theta_1}$
- $\omega_2 = \frac{s\sqrt{1+\theta_1^2}}{\theta_2}$
- $\lambda_1 = \frac{\theta_2}{\theta_1}\sqrt{1+\theta_1^2+\theta_2^2}$
- $\lambda_2 = \frac{\theta_1}{\theta_2}\sqrt{1+\theta_1^2+\theta_2^2}$

If $\sigma_2^2 \to 0$, we obtain $\theta_1 = \frac{\sigma_1}{\sigma_u}$; $\theta_1 = 0$; $s = \sqrt{\sigma_u^2+\sigma_1^2}$; $\frac{1}{\omega_1} = \frac{\sigma_1}{\sigma_u\sqrt{\sigma_u^2+\sigma_1^2}}$. It follows that:

$$E(v|r) = \frac{\sigma_1^2}{\sigma_u^2+\sigma_1^2}r + \frac{\sigma_1}{\sqrt{\sigma_u^2+\sigma_1^2}}\sigma_u\frac{\phi\left(\frac{\sigma_1}{\sigma_u\sqrt{\sigma_u^2+\sigma_1^2}}r\right)}{\Phi\left(\frac{\sigma_1}{\sigma_u\sqrt{\sigma_u^2+\sigma_1^2}}r\right)}$$

This is same as Kogan et al. (2017) upon noting the $1-\Phi(-x) = \Phi(x)$ & $\phi(x) = \phi(-x)$ and $\delta = \frac{\sigma_1^2}{\sigma_u^2+\sigma_1^2}$:

$$E(v|r) = \delta r + \sqrt{\delta}\sigma_u\frac{\phi\left(\frac{\sqrt{\delta}}{\sigma_u}r\right)}{\Phi\left(\frac{\sqrt{\delta}}{\sigma_u}r\right)}$$

Appendix C
Supplementary empirical results

C1 Supplementary results

Table C1: Patent private value and science: Technological Sector

	Panel A: Private Value (*Sci&Non-Sci (separate)*) and Science				
	(1) All	(2) Life Science	(3) Chemicals	(4) ICT	(5) Other
Science Dummy	0.0423***	0.0313***	0.0456***	0.0289***	0.0615***
	(0.0016)	(0.0055)	(0.0042)	(0.0019)	(0.0039)
Constant	15.8473***	16.1388***	16.2184***	15.7511***	15.7988***
	(0.0007)	(0.0035)	(0.0021)	(0.0010)	(0.0012)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
4-digit IPC Fixed Effects	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes
R^2	0.687	0.798	0.736	0.690	0.677
N	1,446,634	108,901	171,121	678,422	411,544
	Panel B: Private Value (KPSS) and Science				
	(1) All	(2) Life Science	(3) Chemicals	(4) ICT	(5) Other
Science Dummy	0.0080***	-0.0276***	-0.0037	0.0049**	0.0153***
	(0.0016)	(0.0056)	(0.0042)	(0.0020)	(0.0039)
Constant	15.8182***	16.1264***	16.2021***	15.7047***	15.7913***
	(0.0007)	(0.0035)	(0.0021)	(0.0010)	(0.0012)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
4-digit IPC Fixed Effects	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes
R^2	0.680	0.793	0.727	0.685	0.666
N	1,446,634	108,901	171,121	678,422	411,544

Note: The dependent variable in panel A is the natural log transformation of the private value of patent: *Sci&Non-Sci (separate)*, estimated using separate distribution assumptions for science-based and other patents. The dependent variable in Panel B is *Patent Value: KPSS (const. USD mn)*: the private value of a patent estimated using the KPSS methodology. We defined "science-based patents" as those ranking in the top three quartiles for the number of non-patent literature (NPL) citations within a specific IPC class and year, provided they cite at least one NPL. All other definitions are identical to those in Table 2. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table C2: Patent private value, technical quality and science premium

	KPSS Estimates	Sci&Non-Sci (separate)	Sci&Non-Sci (mixed)
	(1)	(2)	(3)
Science Dummy	-0.0069***	0.0274***	0.1186***
	(0.0016)	(0.0016)	(0.0016)
log(1+ Patent forward cites)	0.0395***	0.0394***	0.0395***
	(0.0006)	(0.0006)	(0.0006)
Constant	15.7496***	15.7786***	15.6950***
	(0.0012)	(0.0012)	(0.0012)
Avg DV	15.820	15.857	15.796
Year Fixed Effects	Yes	Yes	Yes
4-digit IPC Fixed Effects	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes
R^2	0.681	0.688	0.683
N	1,443,253	1,443,253	1,443,253

Note: The dependent variable in column (1) *Patent Value: KPSS (const. USD mn)*: private value of a patent estimated using the KPSS methodology. The dependent variable in column (2) is *Patent Value: Sci&Non-Sci (separate) (const. USD mn)*: patent's private value estimated using separate distribution assumptions for science-based and other patents. However, patent value is equally distributed among all patents granted on the science patent days (when at least one science-based patent is granted on a given grant date). The dependent variable in column (3) *Patent Value: Sci&Non-Sci (mixed) (const. USD mn)*, we differentiate science-based patents from non-science-based patents using our modified formula. Patent forward cites are the total number of forward citations garnered by a patent over five years. The "science-based patents" are defined as those ranking in the top three quartiles for the number of non-patent literature (NPL) citations within a specific IPC class and year, provided they cite at least one NPL. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Appendix D
Simulated data where signal values are drawn from two different distributions

D1 Simulated data where signal values are drawn from two different distributions

We conduct a simulation to compare the performance of the KPSS method with our proposed approach in estimating the difference in average patent values within firms for two types of patents, where the patent values are drawn from two different distributions. In this analysis, we simulate data for 5,000 firms over a period of 40 years, resulting in daily observations (i.e., 365×40 days). For each firm i, we randomly generate values of σ_{ni} from a normal distribution with a mean of 0 and a standard deviation, σ_i, which is a random value between 0 and 1. The variable x_n is sampled from a half-normal distribution $\mathcal{N}^+(0, \sigma_i)$. For S-type patents, we assume that $x_s \sim \mathcal{N}^+(0, 1.1\sigma_i)$, with $\sigma_{si} = 1.1\sigma_i$, while the noise term follows $\epsilon \sim \mathcal{N}(0, 5\sigma_i)$. On 15% of the days, x_n is drawn from $\mathcal{N}^+(0, \sigma_i)$ for N-type patents, and on 10% of the days, x_s is drawn from $\mathcal{N}^+(0, 1.1\sigma_i)$ for S-type patents.

Next, we compute the realized value $r = I(n) \cdot x_n + I(s) \cdot x_s + \epsilon$, where $I(n)$ and $I(s)$ are indicator functions for N-type and S-type patents, respectively, and ϵ represents the noise. Following the data generation, we first calculate $\mathbb{E}[x_j|r_j]$ using the KPSS method. We then estimate the overall signal-to-noise ratio by applying equation 4 to the simulated data. Additionally, we estimate distinct signal-to-noise ratios for N-type and S-type patents using equation 10, and compute $\mathbb{E}[x_j|r_j]$ separately for both patent types.

In our simulated data, the true coefficient for the Science Patent dummy γ_s in equation 10 is $\ln\left(\frac{(1.1)^2+5^2}{5^2}\right) = 0.47$, while the coefficient for the No Science Patent dummy γ_n is $\ln\left(\frac{(1)^2+5^2}{5^2}\right) = 0.39$. In contrast, the estimated coefficient $\hat{\gamma} = 0.0420$ is obtained from the KPSS method (equation 4), and the estimated coefficients $\hat{\gamma_S} = 0.047$ and $\hat{\gamma_n} = 0.038$ for equation 10.

Table D1: Estimated coefficients of signal-to-noise ratio: One distribution (KPSS) vs. two distributions

	KPSS: One Distribution	Modified: Two distribution
	(1)	(2)
Science Patent Dummy		0.0475***
		(0.0010)
Patent Dummy	0.0420***	
	(0.0006)	
No Science Patent Dummy		0.0384***
		(0.0006)
Constant	-0.0319***	-0.0319***
	(0.0001)	(0.0001)
Firm*Year Fixed Effects	Yes	Yes
Return Day	Yes	Yes
R^2	0.443	0.443
N	72,995,000	72,995,000

Note: In column 1, we estimate the singal-to-noise ratio using equation 4. In column 2 we have two different signal dummies based on the distribution type and used the equation similar to equation 10 To simulate data where signal values are drawn from two different distributions, we proceed as follows. We create 5,000 firms and, for each firm, generate daily data for 365×40 days. For each firm, the ε values are drawn from a normal distribution with a mean of 0 and a standard deviation defined as a random number between 0 and 1 for each firm, multiplied by five, denoted as $\mathcal{N}(0, \sigma_i \times 5)$. The patent value x values are then generated as follows: on most days, the signal value $r = 0$. For 15% of the days, r is drawn from a half-normal distribution $\mathcal{N}^+(0, \sigma_i)$, representing nonscience patents. For 10% of the days, r is drawn from a half-normal distribution $\mathcal{N}^+(0, \sigma_i \times 1.1)$, representing science patents. The actual Science Patent dummy coefficient in our simulated data $ln(\frac{(1.1)^2+5^2}{5^2}) = 0.47$ and the No Science Patent dummy coefficient is $ln(\frac{(1)^2+5^2}{5^2}) = 0.39$ Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table D2: Difference in True vs. Estimated patent values: One distribution vs. two distributions

	True value (x_i)		Estimated value of $\mathbb{E}(x_i \mid r_i)$			
			One distribution		Two distribution	
	(1)	(2)	(3)	(4)	(5)	(6)
Science Patent Dummy	0.0401***	0.0399***	0.0008***	0.0006***	0.0449***	0.0447***
	(0.0002)	(0.0002)	(0.0001)	(0.0000)	(0.0001)	(0.0000)
Constant	0.3992***	0.3992***	0.4131***	0.4132***	0.3950***	0.3950***
	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0000)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	No	Yes	No	Yes	No	Yes
R^2	0.002	0.303	0.000	0.960	0.008	0.957
N	18,247,117	18,247,117	18,247,117	18,247,117	18,247,117	18,247,117
Mean DV	0.415	0.415	0.413	0.413	0.413	0.413

Note: In columns 1 and 2, the dependent variable is the simulated values of x. In columns 3 and 4, we estimate $\mathbb{E}(x|r)$ using the KPSS methodology, while in columns 5 and 6, the dependent variable is $\mathbb{E}(x|r)$ estimated using two different signal-to-noise ratios for x_s and x_n. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table D3: Summary Statistics: True and estimated patent value

	Obs.	Mean	Std. Dev.	10%	50%	90%
True value (x_i)	18247117	0.415	0.433	0.029	0.271	1.011
Estimated Value $\mathbb{E}(x_i\|r_i)$ (One distribution)	18247117	0.413	0.241	0.085	0.412	0.740
Estimated Value $\mathbb{E}(x_i\|r_i)$ (two distribution)	18247117	0.413	0.242	0.085	0.410	0.740

Note: The *simulated values* are the actual simulated values of x as described in D1. The *Estimated Value (One Distribution)* represents the estimated values of $\mathbb{E}(x|r)$ following the KPSS methodology, while the *Estimated Value (Two Distributions)* refers to the estimated values of $\mathbb{E}(x|r)$ obtained using separate signal-to-noise ratios for s-type and n-type patents.

Table D4: Correlation: True and estimated patent value

	True value (x_i)	Estimated Value $\mathbb{E}(x_i\|r_i)$	
		One Distribution	Two Distribution
True value (x_i)	1.000		
Estimated Value $\mathbb{E}(x_i\|r_i)$ (One distribution)	0.557	1.000	
Estimated Value $\mathbb{E}(x_i\|r_i)$ (two distribution)	0.559	0.995	1.000

t statistics in parentheses

Note: The *simulated values* are the actual simulated values of x as described in D1. The *Estimated Value (One Distribution)* represents the estimated values of $\mathbb{E}(x|r)$ following the KPSS methodology, while the *Estimated Value (Two Distributions)* refers to the estimated values of $\mathbb{E}(x|r)$ obtained using separate signal-to-noise ratios for s-type and n-type patents.