



Max Planck Institute
for Innovation and Competition

Max Planck Institute for Innovation and Competition Research Paper No. 28-08

Jonathan Federle, Dietmar Harhoff, and Ann-Christin Kreyer

Value Estimates for U.S. Patent Grants

Value Estimates for U.S. Patent Grants

Jonathan Federle, Dietmar Harhoff, and Ann-Christin Kreyer*

April 2026

Abstract

We propose major amendments to the standard approach for retrieving valuations for U.S. patents from stock market data put forth by Kogan et al. (2017). For most firms, especially since the 1980s, grant events involve multiple, simultaneously issued patents. We argue that the number of simultaneously issued patents affects the signal-to-noise ratio used for filtering patent value estimates out of stock returns. We also propose a method for allocating the firm's capital gain from a week's multiple patent grants to individual patents. We show that our revised values exhibit much higher correlations with external value measurements.

JEL Classification: O30, O31, O34, G10, G14, G12.

Keywords: Stock Market, Patent Disclosure, Patent Grant, Patent Value, Market Capitalization.

*Federle: Kiel Institute (email: jonathan.federle@ifw-kiel.de); Harhoff: Max Planck Institute for Innovation and Competition and CEPR (email: dietmar.harhoff@ip.mpg.de); Kreyer: Max Planck Institute for Innovation and Competition (email: ann-christin.kreyer@ip.mpg.de). The authors acknowledge financial support from Deutsche Forschungsgemeinschaft (DFG) under the auspices of Collaborative Research Center (CRC) 190. We appreciate helpful discussions with Matt Marx, Bronwyn Hall, Tim Simcoe, Brian Wright, David Heller, Fabian Gaessler, Bruno Pellegrino, Felix Poege, and Rainer Widmann. We thank KPSS for their generous help in replicating parts of the results from Kogan et al. (2017). The paper profited from comments received in seminars at the University of Zürich, the Zentrum für Europäische Wirtschaftsforschung (ZEW), University of Munich (LMU) and the IP Day 2024 at Boston University.

1 Introduction

Patent systems are the most comprehensive systematic collection of information on technical advances. That makes them well-suited for addressing a broad range of research questions. A key challenge herein lies in the proper valuation of patents—a task, that is typically complex requiring a deep understanding of highly technical documents and profound industry knowledge. Their value is idiosyncratic and depends crucially on the technical contribution made above and beyond prior art, but also on the nature of complementary assets that the patent holder commands (Teece, 1986). Patent values are therefore highly heterogeneous, and their distribution is highly skewed, meaning that a small fraction of patents accounts for a large share of the total value represented by all patents in a firm’s portfolio, in a sector, or in the overall economy (Scherer, 1965; Harhoff, Scherer and Vopel, 1998).

Abstracting from these technicalities, several studies—most prominently the seminal work of Kogan et al. (2017, henceforth KPSS)—have pioneered to back out market-implied patent values from stock returns around patent grants. They compute the value of individual patents as the average patent-related capital gain per granted patent in a given week to the focal firm. Undoubtedly, the method of Kogan et al. (2017) is a groundbreaking contribution, but only a few studies explicitly consider possible weaknesses of KPSS estimates and guard against them empirically.¹

We expand on the work of KPSS, arguing that the original estimates are subject to at least two² biases. We propose a fundamental generalization of their framework to accommodate these biases and provide a novel corrected dataset. The correlation of our estimates with external value proxies is about 50% higher than that of KPSS values.

First, we change a basic assumption regarding the signal-to-noise ratio (SNR).

¹Higham, de Rassenfosse and Jaffe (2021) comment on the surprisingly low correlation of the KPSS estimates with a large set of other variables that have been shown to have robust statistical association with patent value. Wang (2023) estimates the signal-to-noise ratio separately for different types of patents, but restricts attention to singleton patent issues.

²We defer the discussion of a third issue to a later version of this paper, but acknowledge its presence and implications here. Kogan et al. (2017) assume that the patent, as it will or will not be granted, is publicly known, and that the patent issue only resolves uncertainty regarding the grant decision. This assumption may be justified for a significant portion of the patents covered in their data, in particular when the application has been disclosed before. However, some patents have been kept secret. In an extension of this paper, we identify all patents that are likely still unknown when granted by the USPTO and provide a correction of our estimates to account for this fact. Our revised estimates indicate that none of the results presented here will change qualitatively.

This variable denotes the fraction of return variance around patent grant days related to the patent itself. In the original KPSS framework, the SNR is considered constant over firms and over time. This is problematic for two reasons: (1) Even in a patentless world, firms significantly differ with respect to their return volatility. As such, a firm's volatility is bound to affect the value of its patents, and (2) it is common for multiple patents to be granted to the same firm on the same day. In this case, the signal-to-noise ratio is bound to rise. We allow the signal-to-noise ratio to vary both in the patent-related and the patent-unrelated return components at the firm-day level. Our revised SNR yields patent values that exhibit substantially higher correlations with external patent-value correlates.

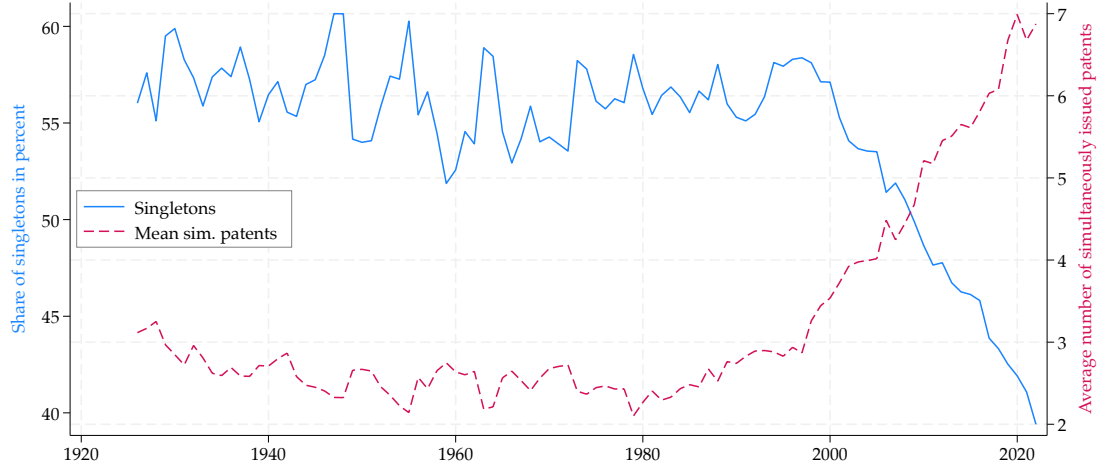
Second, we introduce another degree of freedom for patents that are granted simultaneously with other patents to the same firm on the same day. Naturally, the heterogeneity of patent values that are granted on the same day to the same firm is masked in daily returns. KPSS address this issue by averaging the observed patent-related market value gains over the number of simultaneously granted patents, thus possibly introducing measurement error in about 85% of the patents in their sample. We demonstrate below that averaging is internally inconsistent with important aspects of their data. The values of singleton grants are systematically related to patent characteristics, and the aggregate value of patents issued simultaneously is similarly related to the aggregates of these patent characteristics. Thus, the empirical evidence contradicts an averaging approach. Lastly, the patent economics literature has long argued that patent values follow a highly skewed distribution, which is most likely altered in an averaging operation.

To address the averaging problem, we estimate hedonic patent value regressions, which allow us to assign weights to patents within the daily³ bundle of patent grants while keeping the estimated total over simultaneously granted patents constant. The new value estimates show greatly increased correlations with known patent value correlates. Based on our computations, we provide a completely revised dataset for nearly 3 million patent grants at the USPTO.

There is considerable heterogeneity over time in the number of grants issued per day to any given firm. Figure 1 shows the share of singleton patent grant events and the average number of patents issued per day over time. Both variables remain stable from 1925 to the mid-1980s, after which the average number of patents issued per day begins to increase and the share of singleton grant events starts to decline. We argue that this trend is directly related to the

³Since the USPTO announces patent grants on every Tuesday, we focus on daily data.

Figure 1: Simultaneously issued patents and singletons over time



Note: Time in years depicted on x-axis. Blue solid line denoted on left y-axis and outlines the share of patents issued as singletons (no other patent issued on the same firm day) in percent. Red dashed line denotes the annual average of the number of simultaneously issued patents on the same firm day.

sharp increase in patenting activities by U.S. firms in the 1980s.⁴ While there may be various explanations for the rise of patenting and litigation, the clear time patterns revealed in Figure 1 point to a problem in the currently used KPSS data, going beyond the threat of a time-invariant size bias. KPSS data have been used to study patent litigation, patent portfolio strategies, and the role of patents for VC-financed companies. Many of these phenomena became more prevalent in the 1980s, coinciding with changes in the number of patents issued per day—a central variable driving the signal-to-noise ratio, which is held constant in the original framework. Consequently, the bias in KPSS data may affect studies of these phenomena in complex ways unless researchers prudently control for its causes. We argue that, rather than relying on controls and multiple robustness exercises, it may be preferable to develop a revised set of estimates for the value of patent grants. We propose an approach for doing so and provide such a dataset.

The remainder of this paper is structured as follows. Section 2 introduces our theoretical framework and its empirical implementation. Section 3 describes the construction and properties of our dataset. We proceed to discuss our revised

⁴Hall and Ziedonis (2001) argue that this development initially started in the semiconductor industry and then spilled over into other industries. It was driven by firms' efforts to build patent portfolios that could be used for litigation, which also became more frequent during these years. Moreover, the 1980s also saw growing importance of the venture capital (VC) financing model in the USA. See Hall (2004) for further exploration of the rise of patents in the 1980s.

value estimates and their properties in Section 4. Section 5 presents a discussion of three important empirical relationships that may be viewed differently when using the original KPSS data or our estimates: we study the nexus between value estimates and citations, the relationship between firm size and other patent-level characteristics and patent value. We briefly conclude in Section 6.

2 Theory and empirical framework

2.1 Basic assumptions

Throughout this section, we describe the empirical framework employed to infer patent values from stock market returns. Our point of departure is similar to the framework of KPSS: The market-adjusted return of firm i on day t , $R_{i,t}$ is given by,

$$R_{i,t} = v_{i,t} + \varepsilon_{i,t}. \quad (2.1)$$

Here, $v_{i,t}$ is the aggregate value of all patents of firm i which have been granted in day t expressed as a fraction of the firm's prior day market capitalization and discounted by the market's pre-grant assessment of the grant being successful. Variable $\varepsilon_{i,t}$ denotes the idiosyncratic component of the firm's return, i.e., the component of returns unrelated to patent grants.

It follows that, in absolute terms, the aggregate value of all patents granted on a firm day is given by

$$\lambda_{i,t} = (1 - \bar{p})^{-1} * E[v_{i,t}|R_{i,t}] * M_{i,t-1}, \quad (2.2)$$

where $M_{i,t-1}$ is the pre-grant market capitalization of firm i and $\bar{p} = 0.56$ is the unconditional probability of a successful patent application throughout the entire sample.⁵ Adjusting for the success probability assumes that the market has already priced the patent value prior to the patent grant. In this regard, we mirror the assumptions of KPSS by assuming constant grant probabilities and complete public knowledge of patent values before the grant.

Backing out the conditional patent values, $E[v_{i,t}|R_{i,t}]$, required to compute (2.2), is contingent on distributional assumptions of both $v_{i,t}$ and $\varepsilon_{i,t}$. Since patent grants are assumed to have strictly positive value⁶, KPSS impose a positively

⁵Wang (2023) relaxes the assumption of a homogeneous grant rate, but applies this generalization only to singleton patent issues.

⁶This assumption is probably justified for the vast majority of cases, assuming that the grant does not bring "bad news", e.g., in the form of major restrictions of the claims of the patent,

truncated normal distribution on them, i.e.,

$$v_{i,t} \sim N^+(0, \sigma_{v,i,t}^2). \quad (2.3)$$

The idiosyncratic return component follows a random walk and is assumed to be normally distributed such that

$$\varepsilon_{i,t} \sim N(0, \sigma_{\varepsilon,i,t}^2). \quad (2.4)$$

Given these distributional features, it follows that

$$E[v_{i,t} | R_{i,t}] = \delta_{i,t} R_{i,t} + \sqrt{\delta_{i,t} \sigma_{\varepsilon,i,t}^2} \frac{\phi\left(-\sqrt{\delta_{i,t}} \frac{R_{i,t}}{\sigma_{\varepsilon,i,t}}\right)}{1 - \Phi\left(-\sqrt{\delta_{i,t}} \frac{R_{i,t}}{\sigma_{\varepsilon,i,t}}\right)}, \quad (2.5)$$

with ϕ and Φ denoting the standard normal pdf and cdf, respectively. $\delta_{i,t}$ is the signal-to-noise ratio (SNR) of observed returns and given by

$$\delta_{i,t} = \frac{\sigma_{v,i,t}^2}{\sigma_{v,i,t}^2 + \sigma_{\varepsilon,i,t}^2}. \quad (2.6)$$

2.2 Signal-to-noise ratio

Our first major deviation of the KPSS framework comes into play when estimating the SNR.⁷ KPSS assume this ratio to be constant across the entire sample. The underlying assumption is that both $\sigma_{v,i,t}^2$ and $\sigma_{\varepsilon,i,t}^2$ vary over firms and time, but only in constant proportions to each other. Yet, the idiosyncratic return of firms is—by definition—independent of patent-related information. This is particularly problematic in the context of multiple patents being granted simultaneously, as we would expect the variance of the aggregate value of all patents granted, $\sigma_{v,i,t'}^2$, to increase in the number of patents granted. However, since the variance of the idiosyncratic return component, $\sigma_{\varepsilon,i,t'}^2$, is independent of patent grants, it cannot systematically vary with the number of granted patents. Thus, a constant SNR mechanically imposes the value variance of one patent to equal that of multiple simultaneously granted patents. Since the variance determines not only the

relative to the claims that the market has expected.

⁷To estimate the SNR, KPSS employ the following equation: $\log(R_{i,d}^2) = \gamma I_{i,d} + c Z_{i,d} + \mu_{i,d} + u_{i,d}$, where $R_{i,d}$ denotes the three-day return of firm i starting on day d , and $I_{i,d}$ is an indicator variable equal to one if at least one patent was issued to firm i on day d . The control variable, $Z_{i,d}$, comprises day-of-week dummies, and $\mu_{i,d}$ denotes interacted firm-year fixed effects. The sample is restricted to firms that have been granted at least one patent over the sample period. The SNR is then recovered as $\hat{\delta} = (1 - e^{-\gamma})$. For details, see (Kogan et al., 2017, p.679)

dispersion but also the mean of the positively truncated distribution defining the value of patents in equation (2.3), failure to account for a varying SNR invariably leads to biased patent valuations.

To illustrate the issue, consider the following scenario: A representative example of a patent-unrelated risk captured by the idiosyncratic return variance, $\sigma_{\varepsilon,i,t}^2$, would be the headquarters of firm i to catch fire in day t . Imposing $\sigma_{\varepsilon,i,t}^2$ to vary in constant proportions with $\sigma_{v,i,t}^2$ and acknowledging that the $\sigma_{v,i,t}^2$ should vary with the number of granted patents, effectively implies the probability of patent-unrelated risks like the headquarters catching fire to increase in constant proportions with the number of patents granted by the USPTO on a given day. Since this is arguably implausible, a constant SNR implies that patent values are mechanically biased depending on the number of other patents simultaneously issued.

Instead of imposing a constant SNR across all firm-day observations, we back out both the patent-unrelated return variance and the patent-related variance and let them vary at the firm level. Leveraging the fact that the idiosyncratic return component is by definition independent of patent-related returns, we can compute the measurement error $\sigma_{\varepsilon,i,t}^2$ as the realized variance of returns on non-grant days of firm i in the year preceding patent grant day t .⁸

We now proceed with the definition of the patent-related variance of returns. Recall from equation (2.1) that, on each firm-day, the overall return of a firm is given by the portion of patent-related and patent-unrelated returns. Thus, the total variance of returns is characterized by the following formula:

$$\sigma_{i,t}^2 = \text{Var}(v_{i,t}) + \underbrace{\text{Var}(\varepsilon_{i,t})}_{=\sigma_{\varepsilon,i,t}^2} + 2 \times \underbrace{\text{Cov}(v_{i,t}, \varepsilon_{i,t})}_{=0} \quad (2.7)$$

Thus, the total return variance of a firm on day t equals the sum of the variances of its patent-related returns, its idiosyncratic returns, and their covariance. However, the covariation between patent-related returns and the idiosyncratic return—by definition—must be equal to zero. Since the distribution of patent values follows

⁸In contrast, KPSS back out the variance of the measurement error using their estimate of the SNR through $\sigma_{\varepsilon,i,t}^2 = 3\sigma_{i,t}(1 + 3d_{i,t}\hat{\delta})^{-1}$. Here, $\sigma_{i,t}$ is the volatility of firm i in year t and $d_{i,t}$ is the fraction of trading days on which patents have been granted. Their approach of estimating the measurement error is not feasible when allowing the SNR to differ across the number of patents granted. We provide robustness tests reproducing the original KPSS estimator using our measurement error and show that the resulting patent values are almost identical to the original specification with the Pearson correlation exceeding 0.99.

a positively truncated normal distribution, its variance corresponds to:

$$\text{Var}(v_{i,t}) = \sigma_{v,i,t}^2(1 - 2/\pi). \quad (2.8)$$

Equipped with these inputs, we can back out the observed patent value variance for any day on which a patent has been granted using Equation (2.7). Thus, we define the observed value variance of patents granted to firm i on day t as the difference between the variance of the return of firm i on day t and the variance of the patent-unrelated return, which we defined above as the variance of the returns of firm i on non-grant days in the year preceding day t .

While this non-parametric estimate has intuitive properties and yields, on average, an unbiased estimate of the patent value variance, it also has limitations. Most importantly, it can take on negative values if the return variance on a given day is unusually low compared to the reference period (the year prior to the patent grant).

To address this issue, we estimate how the variance of patent values depends on the number of granted patents. To this end, we introduce the following linear model

$$\sigma_{v,i,t}^2 = \alpha_i + \beta \frac{\#Patents}{M_{i,t-1}} + u_{i,t}, \quad (2.9)$$

which relates the observed value variances of patents granted to firm i on day t to a set of firm fixed effects and the ratio of the number of patents granted to the pre-grant market value of the firm.

In relating the variance of patents not only to the number of patents but to the number of patents relative to the firm's market capitalization, we seek to account for single patents having a larger impact on the values of small firms than on big corporations—at least in *relative* terms. Upon estimating this model, we find that the estimate of β is highly significantly positive ($p\text{-value} < 0.001$).

We then proceed to predict the patent value variances, based on our estimated model above, such that:

$$\hat{\sigma}_{v,i,t}^2 = \hat{\beta} \frac{\#Patents}{M_{i,t-1}}. \quad (2.10)$$

In this way, we arrive at our estimate for the firm-day specific variance of the patent-related return variance.

Having both estimates for the patent-related variance ($\hat{\sigma}_{v,i,t}^2$) and our non-parametric estimate for the variance of returns unrelated to patents ($\hat{\varepsilon}_{v,i,t}^2$), we proceed to plug these values into the formula for the SNR as specified in (2.6).

Equipped with these inputs, we compute the aggregate value of all patents granted to firm i on day t , as given by equation (2.5).

2.3 Simultaneous patent grants

At this juncture, KPSS average the values of all simultaneously granted patents in a given firm week. In this way, the value estimates forfeit any heterogeneity across simultaneously issued patents. Given that both the number of patents granted and the value of patents correlate with firm size, the averaging procedure likely disturbs the patent distribution not only at the sample level but also across firms and time. Moreover, while the mean value may be a reasonable estimator if no additional information about patent value is available, such an assumption is internally inconsistent with the KPSS estimates (and would be inconsistent with our estimates using an amended computation of the SNR). We will show below that the value estimates for singleton patents vary systematically with characteristics of the patents that are known at the grant event (or even earlier), and that this also holds for aggregate relationships. This is not a novel insight, and a large literature has studied such correlations to predict patent values, either for aggregate portfolios or single patents (Harhoff, Scherer and Vopel, 1998; Higham, de Rassenfosse and Jaffe, 2021; Hall, Jaffe and Trajtenberg, 2005).

We seek to overcome the problem of computing patent values for individual patents by regressing patent value estimates on a set of patent quality characteristics publicly known at the time of the grant. We then use these estimates to distribute realized capital gains across simultaneously issued patents according to their estimated value. In doing so, we maintain the total value estimate for the bundle of patents issued. Our reweighing of patent value estimates is therefore mean-preserving (over patents simultaneously granted to the firm). Formally, let $J_{i,t}$ denote the set of patents granted to firm i on day t . Furthermore, let Q denote a set of patent-related quality characteristics. To back out the general relationship between each characteristic and patent values around grants, we estimate the following linear regression

$$\lambda_{i,t} = \alpha_i + \beta_t + \sum_{q \in Q} \left(\psi_q \sum_{j \in J_{i,t}} q_j \right) + u_{i,t}, \quad (2.11)$$

where α_i and β_t denote a set of firm and time fixed-effects, respectively, and $u_{i,t}$ denotes the error term. In this way, ψ_q , yield estimates of how much each patent quality characteristic $q \in Q$ contributes to realized capital gains. Based on these

estimates, we define the predicted value of each patent j as⁹

$$\hat{\eta}_j = \sum_{q \in Q} \hat{\psi}_q q_j. \quad (2.12)$$

Lastly, we use $\hat{\eta}_j$ to disaggregate the firm-day patent values from equation (2.2) to the patent level. Thus, we define

$$\tilde{\zeta}_j = \frac{\hat{\eta}_j}{\sum_{p \in J_{i,t}} \hat{\eta}_p} \lambda_{i,t}, \quad (2.13)$$

where j_i refers to the issuing firm and j_t refers to the grant day of patent j . In this way, we step back from KPSS' original methodology of distributing capital gains in equal proportions to simultaneously granted patents but instead distribute them in a qualitatively meaningful way using a set of publicly known quality characteristics.

3 Data

We base our analysis on patent value data from KPSS, utilizing the most recent dataset available through their public GitHub repository.¹⁰ This dataset, an extension of the original data used in their 2017 QJE paper, includes 3,160,453 patents issued by the USPTO from January 5, 1926, to December 29, 2022. The KPSS dataset associates each patent with a unique permanent firm identifier (permco) provided by the Center for Research in Security Prices (CRSP). Leveraging this identifier, we merge the patent data with daily stock market data for 8,432 U.S. firms between 1926 and 2022 from the Center for Research in Security Prices (CRSP).¹¹ This joint coverage of stock and patent data yields a final dataset of 3,155,528 patents. Hence, we can compute patent value estimates along the lines of KPSS with our SNR amendment for almost the full KPSS 2022 sample (99.6%).

Moreover, we obtain a broad range of patent characteristics as part of our hedonic value estimation from PATSTAT.¹² These characteristics include the cumulative GDP covered by all patent filings based on the priority filing of the

⁹Note that $\hat{\eta}_j$ does not include fixed effects and therefore has no qualitative interpretation in absolute but only in relative terms.

¹⁰See <https://github.com/KPSS2017>

¹¹We utilize variables from CRSP such as daily closing price, stock return, value-weighted market index, and shares outstanding. From the closing price and shares outstanding, we compute each firm's daily market capitalization.

¹²PATSTAT database, version Autumn 2023

Table 1: Summary statistics of patent characteristics

	N	Mean	Median	SD	min	max
<i>Panel A: characteristics of all patents</i>						
#patents	2,703,605	23.99	10	38.68	1.00	637
GDP (patent family)	2,703,605	20.25	17.55	10.78	5.24	76.13
claims (grant)	2,703,605	16.42	16.00	11.15	1.00	868
PL refs (wgt. 3yr cit.)	2,703,605	304.97	55.00	1,167.25	0.00	32,715
NPL refs	2,703,605	5.50	0.00	23.69	0.00	3,230
areas cited	2,703,605	3.62	3.00	3.05	0.00	35
<i>Panel B: patent characteristics in weeks with one grant per firm (singletons)</i>						
#patents	354,670	1.00	1.00	0.00	1.00	1.00
GDP (patent family)	354,670	16.89	14.96	10.19	5.24	76.13
claims (grant)	354,670	16.84	15.00	13.31	1.00	683
PL refs (wgt. 3yr cit.)	354,670	278.10	27.00	1185.00	0.00	32,318
NPL refs	354,670	7.33	0.00	28.06	0.00	2,372
areas cited	354,670	3.91	3.00	3.47	0.00	35
<i>Panel C: patent characteristics in weeks with multiple grants</i>						
#patents	2,348,935	27.47	13.00	40.38	2.00	637
GDP (patent family)	2,348,935	20.76	17.87	10.77	5.24	74.51
claims (grant)	2,348,935	16.36	16.00	10.78	1.00	868
PL refs (wgt. 3yr cit.)	2,348,935	309.02	60.00	1,164.49	0.00	32,715
NPL refs	2,348,935	5.22	1.00	22.95	0.00	3,230
areas cited	2,348,935	3.58	3.00	2.98	0.00	35
<i>Panel D: aggregate characteristics for firm week events</i>						
#patents	678,804	3.99	1.00	8.94	1.00	637
GDP (patent family)	678,804	80.66	23.56	223.75	5.41	14,752.77
claims (grant)	678,804	65.42	24.00	153.06	2.00	10,850
PL refs (wgt. 3yr cit.)	678,804	1,214.58	73.00	4,977.08	0.00	226,974
NPL refs	678,804	21.91	1.00	91.32	0.00	5,979
areas cited	678,804	14.42	5.00	33.25	0.00	2,239

Note: The table shows summary statistics of patent characteristics. #patents are the number of patents that are issued per day. *GDP (patent family)* is the cumulative GDP covered by all patent filings based on the priority filing of the focal US patent. *claims (grant)* are the number of claims in the grant document. *PL refs* are the number of U.S. patent literature references weighted by three-year citations, and *NPL refs* are the number of non-patent literature (NPL) references. *areas cited* are the number of different technical areas cited by the focal patent.

focal US patent (GDP-weighted size of the international patent family)¹³, U.S. patent literature references weighted by three-year citations, claims in the grant document, non-patent literature (NPL) references, and the technical areas¹⁴ cited in the grant document. Since some of the value correlates are only available for patents granted after 1970, our hedonic analysis is currently restricted to 2,703,605 patents issued between 1970 and 2022.

The summary statistics for patent characteristics of these 2,703,605 patents are presented in Table 1, Panel A. On average, each patent is issued together with 24 other patents. The patent families associated with the issued patents cover, on average, a GDP of US\$ (1982) 20.3 trillion. The issued patents typically contain 16 claims and 305 weighted patent literature references (weighted by 3-year citations). On average, the issued patents have 5.50 non-patent literature references. Additionally, they cite an average of 3.62 different technological areas in their backward citations.

We also characterize the singleton observations, defined as single patent grants on a given day, and the non-singleton observations separately (Table 1, Panel B and C). A comparison of the panels points to some selectivity: singleton patent families cover countries with less GDP (16.89 vs. 20.76 US\$ trillion), have fewer patent references (278.10 vs. 309.02), but a larger number of non-patent literature references, i.e., are closer to science (7.33 vs. 5.22). Moreover, singleton patents cite prior art from more technical areas more frequently than non-singletons (3.91 vs. 3.58). These differences are statistically highly significant.

Finally, in Panel D, we provide descriptive statistics at the firm-day level, comprising 678,804 firm-day observations (Table 1, Panel D). On average, firms receive 4 patents per day, with a maximum of 637 patents granted in a single firm-day. We will use these data in our hedonic estimates of aggregated patent bundles issued to firms.

4 Results

In this section, we establish and contrast the results from estimating patent values using the KPSS methodology and our revised model. The first subsection deals with the effects and impact of modifying the SNR, and the second subsection

¹³We use the DOCDB patent family definition provided by PATSTAT. The GDP data are taken from the Penn World Tables and complemented with information from Jordà, Schularick and Taylor (2017) and Funke, Schularick and Trebesch (2023).

¹⁴We use the technical area classification developed by Schmoch (2008). See https://www.wipo.int/export/sites/www/ipstats/en/docs/wipo_ipc_technology.pdf.

outlines solutions to the averaging problem.

4.1 Fixing the signal-to-noise ratio

The natural starting point revealing the dynamic properties of the SNR is to let its constituents, the patent-unrelated return variance and the patent-related variance, vary at the firm-day level (see Section 2.2).

The latter is obtained in estimating regression 2.9, where we regress the observed patent-related variance of returns on the number of patents granted relative to the firms' market capitalization. The results are depicted in Table 2. As is apparent from the table, the variance of the patent-related return components strongly varies with the number of patents granted relative to the grantee's pre-grant market value. The association is highly significant (p -value < 0.01).

Table 2: Estimating the patent-related variance

	$\sigma_{v,i,t}^2$
$\frac{\#patents}{MV_{i,t-1}}$	0.00351*** (0.000629)
FXT	YES
Adj. R^2	0.062
N	41,699,393

Note: Table shows results of estimating regression (2.9). It shows how the variance of the patent-related return component depends on the relative number of patents granted to firms. Both the regressor and the dependent variable are winsorized at the 0.1% and 99.9% levels. FXT: firm-year fixed effects. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The value of the coefficient extends beyond mere descriptives as it allows us to compute revised SNRs based on the number of patents granted relative to the firms' market values.

To illustrate the impact of assuming a constant SNR, consider Table 3, which shows the average SNR in the KPSS and our revised specification. The former assumes a constant SNR, whereas the latter implies a SNR that varies at the firm-day level. As becomes apparent from the table, a fixed SNR masks substantial heterogeneity in the number of patents simultaneously granted.¹⁵

Taking into account the differing SNRs also affects the patent value properties of the sample at large, as shown in Table 4. Since the distribution of the number

¹⁵However, it is important to note that a direct comparison between the KPSS SNR and our revised version is inherently limited, as KPSS analyze three-day returns, whereas our specification focuses on grant-day returns.

Table 3: SNR and number of simultaneously granted patents

#patents	Signal-to-noise ratio	
	KPSS	FHK
1	0.0145	0.0649
5	0.0145	0.08
10	0.0145	0.0873
20	0.0145	0.1331
50	0.0145	0.1442

Note: The table shows the number of simultaneously granted patents against the estimated SNR for the original specification of KPSS and our revised specification.

of simultaneously granted patents carries a substantial right tail, the KPSS specification yields substantially higher values for the majority of patents in the overall sample. Our explanation is that the KPSS estimate overstates patent values for the most frequent observations in the sample, i.e., for event windows with only one or few patent grants. In fact, a static SNR yields mean patent values almost four times higher and median patent values that are about 50% higher than those resulting from our revised SNR. This may help to resolve a puzzle that KPSS themselves note in their paper (see p. 682) where they state that in comparison to estimates by Giuri et al. (2007) their average value of patent estimates “seems a bit high”.

Table 4: Patent value statistics by SNR specification

SNR	Mean	Median	SD	SD / Mean	Skewness	Maximum
KPSS	10.81	3.25	30.01	2.78	14.96	2,683.24
FHK	2.95	2.14	3.03	1.02	2.42	65.24

Note: N=3,158,981. All values in 1982 \$US million. Expressed in 2025 \$US million, the mean of KPSS’s value estimates is 41.94, compared to 11.45 for the FHK estimates. The maximum value in the KPSS estimates is 10,410.97, whereas in the FHK estimates, it is 253.13.

Higham, de Rassenfosse and Jaffe (2021) observe that KPSS values exhibit low correlations with observable patent characteristics. However, as shown in Table 5, allowing the SNR to vary dynamically by firm and day significantly improves these correlations. On average, accounting for differing SNRs increases correlation values at the aggregate level by 50% compared to the original KPSS values.

Table 5: Correlation between capital gains and cumulative value correlates

Patent characteristic	KPSS	FHK
#patents	0.277	0.617
GDP (patent family)	0.263	0.541
PL references (wgt. by 3yr cites)	0.399	0.571
claims (grant)	0.332	0.648
NPL references	0.407	0.552
US cites 3yrs	0.357	0.589
US cites 10yrs	0.332	0.577

Note: The table shows the correlation between value estimates (either the KPSS original or our SNR-corrected FHK estimates) and a number of patent characteristics as summarized in Table 1, Panel D. Patent characteristics are aggregated over simultaneously issued patents. The patent characteristics comprise the cumulative GDP covered by the international patent family (*GDP (patent family)*), the number of U.S. patent literature references weighted by three-year citations (*PL refs*), the number of claims in the grant document (*claims (grant)*), the number of non-patent literature (NPL) references (*NPL refs*), the number of different technical areas cited by the focal patent (*areas cited*), and the number of USPTO citations within 3 years and 10 years of filing (*US cites 3yrs* and *US cites 10yrs* respectively). N=678,743.

In addition, we regress the value estimates from KPSS and our approach on aggregated value correlates to highlight differences between the static and dynamic SNR. Table 6 reports the R^2 and within- R^2 of these regressions. To distinguish the explanatory power of the number of simultaneously granted patents from other correlates, we first regress the estimates on the number of patents alone and then include the additional correlates. The low within- R^2 in column (3) suggests that in the KPSS specification, nearly all variation is explained by firm-year fixed effects. In contrast, in our dynamic SNR-based specification, the within- R^2 remains substantial even after including firm-year fixed effects.¹⁶

4.2 Fixing the averaging problem

To address the averaging issue, we consider different empirical approaches. All of these can be considered hedonic estimates as we model the value of the patent as a function of observable characteristics. We limit our analysis to information known at the time of the patent grant at the USPTO. These characteristics have already been described in section 3. We compute four different hedonic

¹⁶The results are similar if we winsorize the dependent variables at the 0.1% and 99.0% levels (see Table A.1).

Table 6: Regression analyses between capital gains and cumulative value correlates

	KPSS			FHK		
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	22.31*** (0.156)	18.72*** (0.156)	40.66*** (0.0472)	4.929*** (0.0215)	1.460*** (0.0194)	5.970*** (0.00858)
Adj. R^2	0.0799	0.241	0.966	0.345	0.560	0.957
Within R^2	0.0799	0.241	0.000759	0.345	0.560	0.461
N	855,618	855,618	833,641	855,614	855,614	833,637
#patents	YES	YES	YES	YES	YES	YES
Value correlates	NO	YES	YES	NO	YES	YES
FXT	NO	NO	YES	NO	NO	YES

Note: The dependent variables are the original KPSS capital gains and our SNR-corrected FHK estimates. Standard errors in parentheses. Patent characteristics are aggregated over simultaneously issued patents. The patent characteristics comprise the cumulative GDP covered by the international patent family, the number of U.S. patent literature references weighted by three-year citations, the number of claims in the grant document, the number of non-patent literature (NPL) references, the number of different technical areas cited by the focal patent, and the number of USPTO citations within 3 years and 10 years of filing. We also include the quadratic transformations of these patent characteristics. FXT: firm-year fixed effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0$.

value estimates in order to study their properties. Our exercise is motivated by the observation that the original KPSS estimates appear to be only weakly correlated with known patent value proxies. Averaging for a large share of the sample (85%) may well contribute to these low correlations. Naturally, a reweighting using hedonic regressions will strengthen correlations with the variables used as hedonic characteristics and predictors. However, as we show below, the correlation with ex post citation measures—which are not included in the hedonic exercise—also increases. The following regressions use estimates of the total market value gain to firms in a given three-day event window. We apply our SNR correction as described in section 2.2.

Our four estimators are as follows (we relegate regression results to the appendix). The first estimation, *HED-SNG*, focuses on singleton observations and models patent values as a function of the five quality characteristics, firm fixed effects, as well as technical area and grant year fixed effects. The regression results

of this specification are reported in Table A.2, based on 354,633 observations. As expected, we find statistically highly significant coefficients for most of the variables. Since we are mostly interested in obtaining a reasonable prediction of patent value, we refrain from an in-depth interpretation of the coefficients. The fit of the regression is not particularly good in the specification without fixed effects, but improves with the inclusion of year and technical area dummies. Most of the variation is explained by firm fixed effects. We use the results from column (3) for obtaining predictions for all patents and use these to reweigh the averaged patent value estimates such that the firm-day estimate of the total gain in market value is preserved. Note that by definition, the estimated values for singleton grants are maintained here (as in the following models), the rescaling only affects patents granted simultaneously.

Our second model, *HED-AGG*, follows the aggregation logic outlined in equation (2.11) and in the rescaling equation (2.13). We report the results of the regressions in Table A.3. Given the necessity of an additive specification, we use quadratic transformations of our aggregate hedonic characteristics to allow for deviations from a strictly linear relationship. In the absence of aggregation bias, we can use the coefficient estimates to predict pseudo-values at the patent level. We use these as weights to improve upon the averaged values. Since using the coefficients to predict at the individual patent level results in negative predictions for some cases, we impose a value of 0.05 (50,000\$US) in these cases. Our third model, *HED-TOP*, is based on the observation that the value of patent portfolios is strongly determined by their most valuable patents (Scherer and Harhoff, 2000). To estimate the relationship at the aggregate level, we encode all patent characteristics as order statistics. Specifically, we define our regressors as the number of cases, within a given firm-day event, in which an individual patent's characteristics fall into the top five percent and the upper quartile (but not the top five percent) of the distribution. The distribution is computed within technical area and over five-year intervals. Similarly to the previous case, where we used aggregates of the value correlates, predicting values for individual patents results in a few instances of negative predictions, which we handle as previously described.¹⁷ The regression results are presented in Table A.4.

Finally, we use external value information for 3,874 patents, for which we have direct survey estimates from the PatVal project described in Torrisi et al. (2016). Using a log-linear specification, we derive coefficients, then compute

¹⁷We replace negative values with 0.05 in 50.42% cases of the predicted estimates in the *HED-AGG* specification, and for 2% in the *HED-TOP* specification.

predictions for our full sample and use these to reweigh the contributions of these patents to the total market value gain observed within a particular event window (*HED-PATVAL*). The regression results are presented in (Table A.5).

The value estimates resulting from the four models are summarized in Table 7.¹⁸ The rescaling applied does not alter the mean of the distributions. Hence, we preserve the property that, across all estimates, the mean of our value estimates is, on average, 87% lower than the mean of the KPSS values. The median and skewness values vary considerably across the four models. In particular, the *HED-AGG* and *HED-PATVAL* estimators exhibit considerably greater skewness and higher maxima compared to both the original KPSS estimators and our SNR-adjusted estimators with averaging (*FHK*).¹⁹

Table 7: Descriptive statistics for value estimates

	Mean	Median	SD	Skewness	p99	Max.
KPSS	11.89	3.66	32.15	14.11	128.51	2,683.24
FHK	2.61	1.96	2.58	2.16	12.01	53.78
HED-SNG	2.61	1.93	2.64	2.25	12.46	53.7
HED-AGG	2.61	0.89	5.02	7.63	22.73	317.57
HED-TOP	2.61	1.76	2.85	2.33	13.27	53.78
HED-PATVAL	2.61	1.52	3.55	5.06	16.84	221.64

Note: N= 2,703,605. All value estimates in 1982 \$US million. *HED-SNG* are the values predicted using the log-specification with singletons only. *HED-AGG* are the prediction results based on the linear specification for aggregated patent grants. *HED-TOP* are the predicted values of the specifications for aggregates using upper percentile dummies. *HED-PATVAL* are the prediction results using the PatVal survey data.

5 Applications

5.1 Citations

As a first application we explore the relationship between citations and patent value estimates. Computing average values for simultaneously granted patents risks introducing measurement error—high value patents granted at the same

¹⁸Note that we consistently utilize specifications without firm fixed effects.

¹⁹To put these numbers into perspective, in 2025 \$US million, the mean of the hedonic specifications is 10.13. The maximum values for *HED-AGG* and *HED-PATVAL* are 1,232.17 and 859.96, respectively. In contrast, the KPSS mean in 2025 \$US million is 42.25, with a maximum value of 10,410.97.

time as low-value ones would be underestimated, while the value of lesser patents would be overstated. We are concerned that averaging introduces biases, resulting in a low correlation between value correlates and the KPSS value estimates. In other words, relying on value estimates with these biases is likely to distort estimation results in multivariate settings.

Table 8: Citation analysis using averaging estimators (KPSS and FHK)

	KPSS			FHK		
	(1)	(2)	(3)	(4)	(5)	(6)
ln(1+cites)	0.213*** (0.009)	0.168*** (0.009)	0.003*** (0.001)	0.113*** (0.005)	0.087*** (0.005)	0.001*** (0.000)
<i>N</i>	2,703,450	2,703,450	2,685,324	2,703,446	2,703,446	2,685,320
<i>R</i> ²	0.130	0.235	0.960	0.147	0.249	0.959
Firm size	NO	YES	YES	NO	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
FE	CXT	CXT	CXT, FXT	CXT	CXT	CXT, FXT

Note: The dependent variable is ln(patent value estimate). See KPSS (2017), section III, Table II for their original estimates. Standard errors in parentheses, clustered by grant year. Firm size is approximated using the market value of the firm in the week preceding the grant event. CXT: tech classification x grant year. FXT: firm fixed effects x grant year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The exercise undertaken to address this concern essentially replicates an analysis conducted by KPSS (section III). KPSS presented a graph in which they plot a citation measure against their value estimates, identifying an almost linear relationship (see their Fig. II). But the data underlying this analysis is at the aggregate level, where observations are pooled to the level of percentiles. The patent-level regression analysis described in KPSS's Table II appears to confirm our concerns. Initially, the relationship between forward citations and patent value estimates is (without additional controls) strong and statistically significant, yielding an elasticity of 0.174. Once more controls are successively introduced—such as size, volatility, and fixed effects for firm years and technology years—the coefficient diminishes and is no longer statistically significant in the final specification. We now demonstrate that our value estimates exhibit a different pattern.²⁰

We first perform the same exercise as KPSS using averaged values based on a constant SNR (denoted KPSS in Table 8). This approach replicates the pattern

²⁰The summary statistics are reported in Table A.6

Table 9: Citation analysis using hedonic estimators (*HED-PATVAL* and *HED-AGG*)

	HED-PATVAL			HED-AGG		
	(1)	(2)	(3)	(4)	(5)	(6)
ln(1+cites)	0.239*** (0.007)	0.214*** (0.006)	0.152*** (0.005)	0.266*** (0.008)	0.249*** (0.008)	0.196*** (0.012)
<i>N</i>	2,703,446	2,703,446	2,685,320	2,703,446	2,703,446	2,685,320
<i>R</i> ²	0.187	0.261	0.845	0.146	0.168	0.574
Firm size	NO	YES	YES	NO	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
FE	CXT	CXT	CXT, FXT	CXT	CXT	CXT, FXT

Note: The dependent variable is ln(patent value estimate). See KPSS (2017), section III, Table II for their original estimates. Standard errors in parentheses, clustered by grant year. Firm size is approximated using the market value of the firm in the week preceding the grant event. CXT: tech classification x grant year. FXT: firm fixed effects x grant year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

previously described. When we use averaged values based on the dynamic SNR estimate, the results are qualitatively similar. In contrast, using hedonic estimates (*HED-PATVAL* and *HED-AGG*, see Table 9) yields significantly different results. Even with the most aggressive introduction of fixed effects, the citation coefficient decreases but remains relatively large (0.152 in column (3) and 0.196 in column (6)). The *HED-SNG* coefficient estimates behave similar to the averaging estimators, confirming our view that these may be affected by selection bias and that the rescaling based on singletons introduces only minor changes in the patent-level value estimates. The *HED-TOP* coefficient estimates are almost identical to the *HED-PATVAL* and *HED-AGG* estimates (see Table A.7).

The exercise allows us to conclude that averaging has the expected effect of artificially reducing the partial correlation between citations and patent value measures. Our result stands in contrast to the conclusion of KPSS that citations are better considered a measure of scientific value than of monetary value. Although we do not currently endorse a specific variant of our hedonic estimators, the properties illustrated in Table 7 suggest that these estimators are preferable to the original KPSS estimates.

5.2 Patent Term Extension, Standard Essential Patents, and Firm Size

As a second application, we investigate the relationship between additional patent- and firm-level characteristics and our value estimates, as well as those proposed by KPSS. As we noted in the previous application, the use of average values for simultaneously granted patents in KPSS may introduce measurement errors. Specifically, we assess and compare the correlation between estimated patent values and key factors, including patent term extensions, standard essentiality, and firm size.

To analyze the relationship between patent value and patent term extensions, standard essentiality (SEPs), and firm size we regress the estimated patent values on indicator variables. These indicators are expected to be positively correlated with patent value estimates. We include an indicator variable that equals one if the patent belongs to a patent family that contains at least one SEP.²¹ Information whether a patent's term has been extended by the USPTO is obtained from PATSTAT. To account for firm size, we include a categorical variable capturing the number of employees per firm-year, based on data from COMPUSTAT.

Table 10 reports the regression results, controlling for year, firm fixed effects, and year-technology-class fixed effects. In both our *HED-AGG* and *HED-PATVAL* specifications, patents which received a term extension are highly significantly and positively correlated with patent value. In contrast, the correlation between KPSS estimates and patent term extension is negative and statistically insignificant. Similarly, the relationship between SEPs and patent value varies across specifications. SEPs are significantly positively correlated with patent value in our specifications, whereas the KPSS estimates yield substantially smaller coefficients. The results also underscore the positive correlation between firm size and patent value. In all three specifications, the coefficients increase almost linearly with the number of employees per firm-year. However, the coefficients in the KPSS estimates are much larger compared to our *HED-AGG* and *HED-PATVAL*.

Overall, this analysis confirms the hypothesis that the averaging reduces the partial correlation between patent-level characteristics and patent value estimates.

5.3 Patent Value Examples

We further illustrate the variation in value estimates among simultaneously granted patents with two examples.

²¹This classification is based on Brachtendorf, Gaessler and Harhoff (2023)

Table 10: Patent term extension, SEPs, firm size and value estimates

	KPSS	HED-AGG	HED-PATVAL
	(1)	(2)	(3)
patent term ext.	-0.013 (0.0466)	0.227*** (0.0625)	0.168*** (0.0395)
SEP	0.074** (0.035)	0.218*** (0.040)	0.461*** (0.023)
Missing	1.186*** (0.100)	0.625*** (0.072)	0.673*** (0.070)
251–500 Employees	0.207*** (0.033)	0.176*** (0.023)	0.144*** (0.021)
501–2.500 Employees	0.492*** (0.039)	0.373*** (0.027)	0.346*** (0.026)
2.501–10.000 Employees	0.833*** (0.050)	0.585*** (0.042)	0.560*** (0.039)
10.001–50.000 Employees	1.145*** (0.082)	0.749*** (0.057)	0.768*** (0.056)
50.001–250.000 Employees	0.984*** (0.102)	0.596*** (0.067)	0.670*** (0.066)
>250.000 Employees	1.235*** (0.127)	0.849*** (0.087)	0.954*** (0.084)
Constant	-0.165* (0.083)	-0.979*** (0.056)	-0.529*** (0.055)
N	2,702,158	2,702,154	2,702,154
Adj. R^2	0.866	0.502	0.753
Adj. R^2 FE only	0.864	0.501	0.750
Year FE	YES	YES	YES
Firm FE	YES	YES	YES
CXT	YES	YES	YES

Note: The dependent variable is the natural logarithm of the estimated patent value. The base category for firm size is firms with 250 or fewer employees. The “Adjusted R^2 FE” refers to the adjusted R^2 from regressions including only fixed effects. See Table A.8 for details on the categorical variables. Standard errors in parentheses, clustered by grant year. CXT: tech classification x grant year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

First, on May 3, 2022, Apple was granted 43 patents, including the patent with the highest value in our *HED-PATVAL* specification, estimated at \$US 221.64 million. While KPSS assign the same value of \$US 77.58 million to each of these patents, our *HED-PATVAL* estimates exhibit significantly greater variation, as

shown in the left panel of Figure 2.²² Second, on April 5, 1994, IBM received 41 patents. The histogram in the right panel of Figure 2 demonstrates that, again, our *HED-PATVAL* specification captures meaningful variation across patents, whereas KPSS assigns a uniform value of \$US 3.37 million to all.

These examples highlight the ability of our improved method to better capture heterogeneity in patent value among simultaneously granted patents.

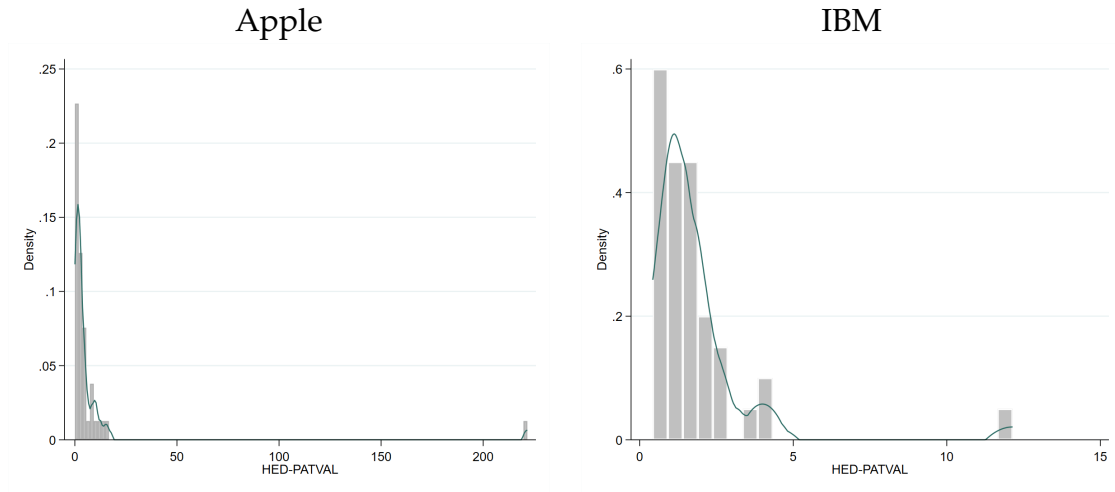


Figure 2: Distribution of patent value estimates

Notes: The left histogram depicts the distribution of patent values, estimated using the *HED-PATVAL* specification, for 43 patents simultaneously granted to Apple on May 3, 2022. In contrast, KPSS assign each of these patents a uniform value of \$77.58 million (deflated to 1982). The right histogram shows the corresponding distribution for 41 patents granted to IBM on April 5, 1994, where KPSS similarly attribute a uniform value of \$3.37 million (deflated to 1982) per patent.

²²All value estimates are in 1982 \$US.

6 Conclusion

In their pioneering work, KPSS introduced a novel method for computing patent value estimates based on short-term stock market reactions to patent grant events. Their approach has sparked renewed interest in employing patent value estimates across various analytical studies. Computing the estimates can be highly standardized, and the resulting data can cover a broad range of sectors and time periods.

In our analysis, we identify two sources of bias in the KPSS value estimates that may impose limitations in subsequent applications. We propose major amendments to their approach to address these issues.

We first argue that assuming a constant SNR introduces a non-trivial bias. Allowing both the patent-related and patent-unrelated return components to vary at the firm-day level, and permitting the SNR to adjust based on the number of simultaneously granted patents, introduces two degrees of freedom that lead to significant changes in value estimates. These adjustments make the estimates more consistent with prior literature and internally coherent. Comparing our dynamic SNR to KPSS's static SNR in relation to the number of simultaneously issued patents further reveals substantial heterogeneity, which the static SNR in KPSS masks. Compared to KPSS estimates, our FHK estimates of patent-level capital gains to firms are considerably lower. The mean of the KPSS patent value distribution is nearly five times higher than ours, while the median is more than 50% higher.

We address the limitations of the averaging approach by using hedonic regressions, which yield weights that allow us to redistribute value among individual patents issued simultaneously to a given firm. This revision improves the correlation between our modified values and established value correlates. Examining the relationship between forward citations and patent value estimates, we also show that our hedonic adjustment significantly affects the citation coefficient in a log-linear value regression, even though forward citations to the focal patent are not included in the hedonic computations. In addition, the hedonic patent values exhibit a stronger correlation with patent value indicators, such as standard essentiality and patent term extensions, compared to KPSS. Furthermore, we find suggestive evidence that the value of patents granted to smaller firms is overestimated in KPSS. Considering the central role of firm size in various empirical studies within innovation research, industrial organization, and related fields, we expect this bias to spill over to other variables, such as

capital investment, R&D expenditures, and IT-related activities.

The contribution of this paper is first one of clarification—despite of the widespread use of the KPSS value data, there have been only very few investigations seeking to understand which assumptions drive the main results of KPSS. We think that our work will enhance the understanding of the approach chosen by KPSS and that it will offer significant improvements. Moreover, we hope that the paper will open the discussion on how stock market reactions can be better exploited for the purpose of economic research on innovation and competition. Towards that objective, we plan to extend our analysis provided here to specific relationships such as the nexus between patent value and R&D inputs, the impact of science and knowledge externalities on patent value, and the influence of competition on stock market assessments of patent value. Finally, the value estimates provided in this paper may have applications that we as authors may not anticipate.

References

- Brachtendorf, Lorenz, Fabian Gaessler, and Dietmar Harhoff.** 2023. "Truly standard-essential patents? A semantics-based analysis." *Journal of Economics & Management Strategy*, 32(1): 132–157.
- Funke, Manuel, Moritz Schularick, and Christoph Trebesch.** 2023. "Populist leaders and the economy." *American Economic Review*, 113(12): 3249–3288.
- Giuri, Paola, Myriam Mariani, Stefano Brusoni, Gustavo Crespi, Dominique Francoz, Alfonso Gambardella, Walter Garcia-Fontes, Aldo Geuna, Raul Gonzales, Dietmar Harhoff, et al.** 2007. "Inventors and invention processes in Europe: Results from the PatVal-EU survey." *Research policy*, 36(8): 1107–1127.
- Hall, Bronwyn H.** 2004. "Exploring the patent explosion." *The Journal of Technology Transfer*, 30: 35–48.
- Hall, Bronwyn H, Adam Jaffe, and Manuel Trajtenberg.** 2005. "Market Value and Patent Citations." *RAND Journal of Economics*, 16–38.
- Hall, Bronwyn H., and Rosemarie Ham Ziedonis.** 2001. "The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979-1995." *The RAND Journal of Economics*, 32(1): 101–128.
- Harhoff, Dietmar, Frederic M Scherer, and Katrin Vopel.** 1998. "Exploring the tail of patent value distributions."
- Higham, Kyle, Gaétan de Rassenfosse, and Adam B. Jaffe.** 2021. "Patent Quality: Towards a Systematic Framework for Analysis and Measurement." *Research Policy*, 50(4): 104215.
- Jordà, Òscar, Moritz Schularick, and Alan M Taylor.** 2017. "Macrofinancial history and the new business cycle facts." *NBER macroeconomics annual*, 31(1): 213–263.
- Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman.** 2017. "Technological Innovation, Resource Allocation, and Growth*." *The Quarterly Journal of Economics*, 132(2): 665–712.
- Scherer, Frederic M.** 1965. "Firm size, market structure, opportunity, and the output of patented inventions." *The American economic review*, 55(5): 1097–1125.
- Scherer, Frederic M, and Dietmar Harhoff.** 2000. "Technology policy for a world of skew-distributed outcomes." *Research policy*, 29(4-5): 559–566.
- Teece, David J.** 1986. "Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy." *Research policy*, 15(6): 285–305.
- Torrìsi, Salvatore, Alfonso Gambardella, Paola Giuri, Dietmar Harhoff, Karin Hoisl, and Myriam Mariani.** 2016. "Used, blocking and sleeping patents: Empirical evidence from a large-scale inventor survey." *Research Policy*, 45(7): 1374–1385.
- Wang, Wenjun.** 2023. "Valuation of Patents using Stock Market Responses." PhD diss. University of California, Berkeley.

Online Appendix to Value Estimates for US Patent Grants

Jonathan Federle, Dietmar Harhoff, and Ann-Christin Kreyer

February 2026

A Additional Tables

A.1 Tables – Fixing the SNR

Table A.1: Winsorized regression analyses between capital gains and cumulative value correlates

	KPSS			FHK		
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	4.769*** (0.0175)	1.080*** (0.0699)	0.0279 (0.0259)	1.618*** (0.00241)	4.704*** (0.00869)	2.388*** (0.00471)
Adj. R^2	0.0799	0.241	0.966	0.345	0.560	0.957
Within R^2	0.0799	0.241	0.000759	0.345	0.560	0.461
N	855,618	855,618	833,641	855,614	855,614	833,637
#patents	YES	YES	YES	YES	YES	YES
Value correlates	NO	YES	YES	NO	YES	YES
FXT	NO	NO	YES	NO	NO	YES

Note: The dependent variables are the original KPSS capital gains and our SNR-corrected FHK estimates. The dependent variables are winsorized at the 0.1% and 99.9% levels. Standard errors in parentheses. Patent characteristics are aggregated over simultaneously issued patents. The patent characteristics comprise the cumulative GDP covered by the international patent family, the number of U.S. patent literature references weighted by three-year citations, the number of claims in the grant document, the number of non-patent literature (NPL) references, the number of different technical areas cited by the focal patent, and the number of USPTO citations within 3 years and 10 years of filing. We also include the quadratic transformations of these patent characteristics. FXT: firm fixed effects x grant year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.2 Tables – Hedonic Regressions

Table A.2: Log-linear specifications for singletons (*HED-SNG*)

independent variables	(1)	(2)	(3)	(4)
GDP (patent family)	-0.021*** (0.004)	0.034*** (0.005)	0.029*** (0.005)	0.027*** (0.002)
PL refs (weighted by 3yr-cites)	0.004*** (0.001)	0.007*** (0.002)	0.022*** (0.002)	0.006*** (0.001)
claims (grant)	-0.020*** (0.003)	0.010*** (0.003)	0.008*** (0.003)	0.002 (0.001)
NPL refs	-0.010*** (0.002)	-0.023*** (0.002)	-0.033*** (0.002)	0.005*** (0.001)
areas cited	-0.064*** (0.004)	-0.047*** (0.004)	-0.064*** (0.004)	-0.001 (0.002)
<i>N</i>	354,633	354,633	354,633	353,324
<i>R</i> ²	0.002	0.030	0.064	0.868
Year FE	NO	YES	YES	YES
Tech. class FE	NO	NO	YES	YES
Firm FE	NO	NO	NO	YES

Note: This table presents the results from estimating equation (2.11), focusing on singleton observations. The dependent variable is the natural logarithm of the patent-related return for firm *i* on day *t*. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.3: Linear specifications for aggregates (*HED-AGG*)

independent variables	(1)	(2)	(3)
#patents	1.287*** (0.011)	5.519*** (0.020)	3.111*** (0.015)
GDP (patent family)	-0.074*** (0.000)	-0.501*** (0.002)	-0.249*** (0.001)
(GDP (patent family)) ²		0.008*** (0.000)	0.004*** (0.000)
PL refs (weighted by 3yr-cites)	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)
(PL refs) ²		-0.000*** (0.000)	-0.000*** (0.000)
claims (grant)	0.074*** (0.000)	0.109*** (0.001)	0.060*** (0.000)
(claims (grant)) ²		-0.001*** (0.000)	-0.000*** (0.000)
NPL refs	0.046*** (0.000)	0.028*** (0.000)	0.003*** (0.000)
(NPL refs) ²		0.000*** (0.000)	0.000*** (0.000)
areas cited	0.100*** (0.002)	0.282*** (0.004)	0.036*** (0.003)
(areas cited) ²		-0.006*** (0.000)	0.003*** (0.000)
<i>N</i>	678,743	678,743	677,438
<i>R</i> ²	0.530	0.583	0.822
Year FE	YES	YES	YES
Firm FE	NO	NO	YES

Note: This table presents the results from estimating equation (2.11), focusing on aggregated values. The dependent variable is the natural logarithm of the patent-related return for firm *i* on day *t*. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.4: Specifications for aggregates using upper percentile coding (*HED-TOP*)

independent variables	(1)	(2)	(3)
TOP5 GDP (patent family)	0.457*** (0.018)	-0.737*** (0.022)	-0.385*** (0.015)
TOP5 PL refs (weighted by 3yr-cites)	3.753*** (0.047)	1.639*** (0.046)	1.324*** (0.031)
TOP5 claims (grant)	3.933*** (0.031)	-1.202*** (0.035)	-0.758*** (0.024)
TOP5 NPL refs	7.115*** (0.042)	1.956*** (0.044)	1.245*** (0.029)
TOP5 areas cited	4.752*** (0.043)	2.277*** (0.040)	2.096*** (0.027)
TOP25 GDP (patent family)		-1.273*** (0.009)	-0.366*** (0.007)
TOP25 PL refs (weighted by 3yr-cites)		1.540*** (0.019)	1.117*** (0.013)
TOP25 claims (grant)		2.695*** (0.012)	1.703*** (0.009)
TOP25 NPL refs		1.377*** (0.014)	0.104*** (0.010)
<i>N</i>	678,743	678,743	677,438
<i>R</i> ²	0.329	0.465	0.785
Year FE	NO	YES	YES
Firm FE	NO	NO	YES

Note: This table presents the results from estimating equation (2.11). The dependent variable is the natural logarithm of the patent-related return for firm *i* on day *t*. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.5: Log-linear specifications using PatVal survey data (*HED-PATVAL*)

independent variables	(1)	(2)	(3)
GDP (patent family)	1.019*** (0.144)	0.801*** (0.145)	0.943*** (0.156)
PL refs (weighted by 3yr-cites)	0.087*** (0.031)	0.225*** (0.035)	0.260*** (0.037)
claims (grant)	0.228*** (0.064)	0.239*** (0.064)	0.235*** (0.064)
NPL refs	0.193*** (0.040)	0.084* (0.043)	0.100** (0.044)
areas cited	0.403*** (0.089)	0.226** (0.093)	0.255*** (0.094)
<i>N</i>	3,874	3,874	3,874
<i>R</i> ²	0.062	0.107	0.118
Year FE	NO	YES	YES
Tech. class FE	NO	NO	YES

Note: This table presents the results from estimating equation (2.11), focusing on aggregated values. The dependent variable is the natural logarithm of the PatVal values. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.3 Tables – Citation Analysis

Table A.6: Summary statistics of forward citations

	Panel A All Patents	Panel B Singletons	Panel C Multiple Grants	Panel D Firm-Week Aggregates
N	2,703,605	354,670	2,348,935	678,804
Mean	22.46	32.22	23.48	29.31
Median	10.00	14.00	10.00	12.00
SD	60.24	74.90	57.62	69.62
min	0.00	0.00	0.00	0.00
max	4,024.00	3,892.00	4,024.00	4,024.00

Note: The table shows summary statistics of the forward citations used in the regressions in section 5.1.

Table A.7: Citation analysis using hedonic value estimates (*HED-TOP* and *HED-SNG*)

	HED-TOP			HED-SNG		
	(1)	(2)	(3)	(4)	(5)	(6)
ln(1+cites)	0.212*** (0.014)	0.185*** (0.013)	0.101*** (0.012)	0.112*** (0.005)	0.086*** (0.005)	0.001*** (0.000)
<i>N</i>	2,703,446	2,703,446	2,685,320	2,703,446	2,703,446	2,685,320
<i>R</i> ²	0.154	0.233	0.814	0.177	0.274	0.957
Firm size	NO	YES	YES	NO	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
FE	CXT	CXT	CXT, FXT	CXT	CXT	CXT, FXT

Note: The dependent variable is ln(patent value estimate). See KPSS (2017), section III, Table II for their original estimates. Standard errors in parentheses, clustered by grant year. Firm size is approximated using the market value of the firm in the week preceding the grant event. CXT: tech classification x grant year. FXT: firm fixed effects x grant year. *** p<0.01, ** p<0.05, * p<0.1.

A.4 Tables – Applications

Table A.8: Distribution of Categorical Variables

	Frequency	Percent	Cumulative
Patent Term Extension			
Patent term extension	364	0.01	0.01
No extension	2,703,241	99.99	100
Standard Essential Patents			
SEP	13,865	0.51	0.51
No SEP	2,689,740	99.49	100
Firm Size Category			
Missing	52,500	1.94	1.94
≤250 Employees	42,723	1.58	3.52
251–500 Employees	29,385	1.09	4.61
501–2,500 Employees	106,994	3.96	8.57
2,501–10,000 Employees	233,074	8.62	17.19
10,001–50,000 Employees	679,023	25.12	42.30
50,001–250,000 Employees	1,148,169	42.47	84.77
>250,000 Employees	411,737	15.23	100.00
<i>N</i>	2,703,605		

Note: