Precision and bias in the assessment of essentiality rates in firms' portfolios of declared SEPs

Justus Baron* Northwestern University

Tim Pohlmann† IPlytics

November 2021

Abstract

A large and increasing number of patents are declared by their owners to be potentially essential to technology standards developed in open Standards Development Organizations (SDO). Policy makers and many stakeholders view the difficulty to identify actual Standard-Essential Patents (SEP) among these declared potential SEPs as a significant concern. Reliable information on the numbers of actual SEPs held by different firms is often seen as important for the determination of fair, reasonable, and non-discriminatory (FRAND) licensing terms for SEP portfolios. We compare the relative merits of three different approaches to the estimation of the numbers of actual SEPs held by different firms: analyses of random samples, predictive modeling using observable patent characteristics, and a more superficial individual examination of every declared SEP. Superficial checks of every declared SEP and predictive modeling may achieve greater precision than simple sampling, but are susceptible to systematic bias. We recommend sampling for the estimation of essentiality ratios in large firm portfolios of declared SEPs; while predictive modeling is useful for the analysis of larger numbers of smaller SEP portfolios. For small portfolio sizes, a light-touch review of all declared SEPs may also be appropriate, provided that the assessment error is generally zero-centered.

JEL-Classification:JEL codes here

Keywords: Standard-Essential Patents (SEP), essentiality checks, sampling, semantic analysis, machine learning

*Research Associate, Searle Center on Law, Regulation, and Economic Growth, Northwestern University Law School, 375 East Chicago Avenue, Chicago, IL 60611. E-mail: justus.baron@law.northwestern.edu.
Research at the Center on Law, Business, and Economics has benefited from financial support from Qualcomm and Intel, among other sponsors.

†CEO and founder, IPlytics. E-mail: pohlmann@iplytics.com

1 Introduction

Standardized Information and Communication Technologies (ICT) are undergoing rapid innovation, and expanding into an increasing number of industries. A whole new spectrum of products and services will be interconnected and interoperable, altering how we interact, work and communicate with the world around us. This development relies on the standardization of innovative - and often patented - technologies in open, consensus-based, standards-development organisations (SDOs). Therefore, there is an increasing number of ICT standards that are subject to standard essential patents (SEPs).[1] SEPs claim inventions that are necessary for any implementation of the standard.

SDOs require companies participating in standards development to disclose any *potential* SEPs, and indicate whether they are willing to license those patents that are essential under fair reasonable and non-discriminatory (FRAND) terms to implementers of the standard. SEP declarations thus primarily aim to ensure that all SEPs are subject to a licensing commitment pursuant with the SDO's patent policy. Increasingly, these declarations of potential SEPs are however also used by implementers of a standard to obtain information on the patents for which they may require a license.

Given the large number of declarations of potential SEPs for some standards, implementers and SEP holders usually negotiate licenses for entire portfolios of SEPs. In some cases, the value of a license is determined by reference to the overall value of the standardized technology, in combination with the share of that value created by the patented technology in the portfolio.[2]

The apportionment of the value of a technology standard to different SEP portfolios is subject to significant controversy. It is generally understood that the technological significance of individual patents is very heterogeneous. Nevertheless, there is no agreement on methods that could reliably indicate the specific value of individual patents. As a practical matter, some form of patent counts is therefore often used in the determination of licensing terms.[3] In this context, policy makers such as the European Commission are calling for greater transparency on the number of patents owned by different firms that are actually essential to technology standards.[4]

[1]Prominent examples of ICT standards subject to numerous SEPs are 4G and 5G cellular communication standards, WiFi (802.11), Bluetooth, near field communication (NFC) and radio frequency identification (RFID).

[2]An example of a court decision relying on such a *top-down* approach to the valuation of SEP licenses is the ruling of the District Court for the Northern District of Illinois in Innovatio (U.S. District Court of the Northern District of Illinois, Eastern Division, In re Innovatio IP Ventures, LLC Patent Litigation, Case No. 11 C 9308). While decisions relying exclusively or primarily on top-down approaches to SEP license valuation remain the exception, a larger number of decisions have used top-down approaches as complements to more traditional approaches (i.e. analyses of comparable licenses), e.g. the UK High Court in Huawei v Unwired Planet (UK High Court of Justice, Unwired Planet International -v- Huawei Technologies and ors, [2017] EWHC 711

[3]In Unwired Planet v. Huawei, Judge Birss found that "There was ample evidence before me that apart from Ericsson [...], parties negotiating SEP licences in fact use methods which are based on patent counting. That is evidence which supports a finding that a FRAND approach to assessing a royalty rate is to engage in some kind of patent counting. Indeed when one thinks about it some sort of patent counting is the only practical approach at least for a portfolio of any size." ([2017] EWHC 711 at 182)

[4]In 2017, the European Commission noted that "Evidence points to the risk of broad over-declarations and makes a strong case for more reliability with respect to SEP essentiality." European Commission,

While databases of declarations of potential SEPs are often used as a starting point, and in some cases stakeholders erroneously use counts of potential SEPs as a measure of actual SEPs owned by different firms, these databases include numerous patents that are not actually essential. Different studies have produced different estimates of the share of actual SEPs among declared SEPs, and have often indicated that these shares may vary significantly between different firms.[5] In complex licensing disputes, the share of actual SEPs among different firms' declared (potential) SEPs has been a particularly thorny issue of contention.[6]

In view of increasing transparency and reducing transaction costs in the licensing of SEPs, different approaches to the assessment of the number of *truly* essential patents in different portfolios of declared SEPs are currently being discussed. One possible approach is a third-party examination of every declared SEP. Bekkers et al. (2020) report the results of a pilot project based on an eight-hour analysis of each declared SEP by patent examiners, and propose generalizing such individual essentiality assessments to the population of declared SEPs. A second approach is to carry out in-depth essentiality assessments within randomly drawn samples of declared SEPs, and to extrapolate the observed essentiality ratios in different firms' portfolios to the larger population. A third approach is to use explanatory variables to predict whether declared SEPs are actually essential (Brachtendorf et al., 2020). The relationship between observable variables and essentiality can be assessed in a random sample of declared SEPs submitted to essentiality checks, and then be used to make predictions about essentiality rates in the general population.

While different studies present different individual approaches to the assessment of essentiality rates in larger populations of declared SEPs, to the best of our knowledge, there currently is no study that compares the merits and downsides of these different approaches. In this paper, we use technical experts' essentiality assessments for a randomly drawn sample of 1,000 patent families declared to be potentially essential to a 5G TS to make different predictions of essentiality rates by firm portfolio in the population of declared 5G SEPs (currently comprising approx. 25,000 patent families).

"Setting out the EU approach to Standard Essential Patents" COM(2017) 712 final. In its more recent "Intellectual Property Action Plan" of 2020, the Commission went one step further: "The Commission will for instance explore the creation of an independent system of third-party essentiality checks in view of improving legal certainty and reducing litigation costs." European Commission, "Making the most of the EU's innovative potential - An intellectual property action plan to support the EU's recovery and resilience", COM(2020) 760 final.

[5]Goodman and Myers (2005) e.g. found that 157 of 732 (21.4%) analzyed patents declared essential to 3GPP were judged to be "probably essential" by independent experts. In January 2010, Fairfield Resources International, Inc. analyzed 1,115 patents declared as essential to 3GPP Release 8 (LTE and SAE). Of these 210 families, 105 families (50%) have at least one patent judged essential or probably essential. Cyber Creative Institute, in June 2013 evaluated 2,129 patents declared essential to LTE, and deemed that 56.0% of the analyzed patents were essential ("The invention contained in the patent matches the standards"), while 29.0% "partially match the standard". Stitzing et al. (2017) report results of a PA consulting analysis; finding an average essentiality rate of 35.2% in a sample of 3,917 US patents declared essential to LTE.

[6]Recent cases in which determinations of essentiality shares played a significant role include the (vacated) decision by the U.S. Disctrict Court of the Central District of California, in TCL v. Ericsson (TCL's expert testified in February 2017 that according to a Concur IP analysis, 37.3% of 2,600 evaluated patent families declared essential to 2G, 3G, and 4G cellular communication technologies were essential); and the Unwired Planet v Huawei at the UK High Court.

We use the sample to estimate a predictive model, using a very comprehensive set of variables, including a semantic similarity score between patents' independent claims and the text of the TS section, as well as other variables (e.g. technical contributions to standards development, attendance in standards meetings, citations between patents as well as between patents and standards documents). As a benchmark, we also use the sample to directly observe essentiality rates in different firms' SEP portfolios, which we then extrapolate to the larger population.

Based on these analyses, we compare the out-of-sample performance of the different approaches. We therefore randomly split our sample of thousand patents in two subsamples of equal size, and use one subsample to predict essentiality rates in the other. Repeating this numerous times with different randomly drawn subsamples, we assess a method's performance along different dimensions, and in particular precision and bias. Precision is the mean absolute value of the prediction error, whereas bias is the mean of the deviation between predicted and actual rates per firm (a method is unbiased if the mean is zero, meaning that the prediction is just as likely to over-as to under-estimate the essentiality rate in an individual firm's portfolio).

Just like any other sampling-based approach, our predictions for essentiality rates in firms' portfolios of declared 5G SEPs hinge on the accuracy of the assessments of the sample patents. Strictly speaking, our analysis does not reveal the performance of different methods to predict "true" essentiality rates of declared SEPs (which are unobserved), but the capacity of different methods to accurately predict what the outcome of a particular assessment would be if carried out in the larger population of patents. As true essentiality (and thus assessment error) is unobserved, we simulate assessment errors to compare the implications of assessment uncertainty for the performance of different methods of extrapolation. We also use appropriately calibrated simulations of assessment errors to compare the performance of higher quality assessments in a smaller sample with lower quality assessments of larger samples or entire populations.

We find that there is a trade-off between precision and bias - while extrapolating essentiality rates from randomly drawn samples may produce significant prediction errors (especially for small portfolios, resulting in very small samples), predictive models and light-touch reviews of every single patent (while sometimes more accurate on average) have the potential to produce systematic bias. For larger portfolios, however, sampling is generally both less susceptible to bias *and* more accurate than the other methods, and should generally be the method of choice. This also implies that policy makers should encourage market participants to focus the limited resources available for essentiality checks on appropriately chosen samples of declared SEPs, rather than requiring or mandating essentiality checks for entire populations of declared SEPs.

While we compare different methods for the determination of the number and share of actual SEPs in different portfolios of declared SEPs, we do not analyze *whether* the number of actual SEPs should be used for the apportionment of FRAND royalties from a reasonable aggregate royalty, or whether such apportionment from an aggregate value of the standard should be undertaken at all. We only compare the *relative* merits of different approaches. There remain significant questions regarding the overall usefulness of patent counting, e.g. whether the value contribution of different SEPs is sufficiently homogeneous for counts of confirmed SEPs to be meaningful; and whether the aggregate value of a standard can be assessed reliably enough for top-down approaches to be viable. We leave these important questions to future research.

The remainder of this paper is organized as follows. Section 2 presents an overview of the different approaches to essentiality assessments, including a discussion of the existing literature.

2 Overview of Different Approaches

In principle, there are at least three families of different approaches to the determination of essentiality rates in patent holders' portfolios of declared SEPs: comprehensive patent-by-patent assessments of all patents, sampling (i.e. in-depth analyses of randomly drawn samples, and extrapolation of observed essentiality rates in the sample to predicted essentiality rates in the population), and predictions using other, more readily observable patent characteristics as variables, which can be used to "train" predictive models, e.g. logistic regression models or tree-based machine learning approaches, in a randomly drawn sample to predict essentiality rates in the population.

2.1 Patent-by-patent assessments of individual patents

Patent-by-patent essentiality analyses of large populations of declared SEPs are currently confined to specific settings, such as patent pool licensing - for antitrust reasons, patent pools generally aim to ensure that only actual SEPs are included in the pool license. Pool licensing administrators recruit independent experts to conduct essentiality checks of potential SEPs submitted for inclusion into the pool licenisng program. The incentive structure and governance of pools is often viewed as conducive to objective examinations.[7] Pools also charge patent holders significant fees for essentiality checks, thus providing technical experts with sizeable resources - Bekkers et al. (2020) report that pools charge between 5,000 and 10,000 Euro per patent for such an assessment, on average.

The cost of applying this relatively thorough standard of examination to every declared SEPs appears prohibitive.[8] In spite of their price tag, these essentiality checks are unlikely to fully resolve uncertainty regarding the actual essentiality of individual declared SEPs, as the essentiality of SEPs checked by pools' experts can be, and relatively frequently is, subject to additional challenges in court.[9]

[7]In its 1999 Business Review Letter of the DVD6C patent pool, the U.S. Department of Justice e.g. stated that "each Licensor will benefit monetarily from the exclusion of other Licensors' non-'essential' patents and accordingly has a strong incentive to encourage the expert to review other Licensors' patents critically, and to bring to the expert's attention any patents that have ceased to be 'essential.'" https://www.justice.gov/sites/default/files/atr/legacy/2012/08/01/2485.pdf

[8]Baron and Pohlmann (2018) identified 139,620 different patents from various SDO databases. The iplytics database currently includes more than 180,000 declared SEPs. There may be large numbers of additional potential SEPs, as many SDOs allow for blanket declarations instead of requiring specific disclosures of individual patents. Applying the lower end of the range of the fees currently charged by pools (5,000 USD) to an estimated number of 200,000 of potential SEPs yields a cost erstimate of 1bn USD for the existing stock of potential SEPs alone.

[9]To the best of our knowledge, there is no empirical evidence on how the frequency and the outcomes of essentiality challenges to pooled SEPs in court differ from those to other declared SEPs.

There are at least two potential mechanisms for reducing the cost of generalized essentiality checks of individual patents: first, one may conduct less costly (but still objective) essentiality assessments. Bekkers et al. (2020) conducted a pilot study of essentiality assessments at an intermediate level of rigor - professional patent examiners were asked to provide an analysis of each declared SEPs in a sample after an analysis of not more than 8 hours.[10] Generalizing such assessments would still involve more than 1 million hours of highly skilled labor for the assessment of the existing stack of SEP declarations, alone.

At the same time, placing time and resource constraints on examiners introduces additional measurement error - Bekkers et al. (2020) find that assessments in their studies were consistent with more comprehensive assessments by pool examiners in approx. 75% of the cases. This incremental measurement error would result in misclassification of approximately 50,000 patents as essential or non-essential (in the existing stock of declared SEPs alone), so that assessments of any individual patents would face challenges and appeals. Whether such assessments would result in more trustworthy estimates of the *overall number* of actual SEPs held by different firms depends on how assessment inconsistencies are distributed between different firms' patent portfolios.

Second, one may apply generalized checks to non-random subsets of the population of declared SEPs. The contribution of the EU Commission's expert group on SEPs e.g. includes proposals to check one patent per family, instead of all declared SEPs (SEPs-Expert-Group, 2021). While lower than the number of declared SEPs, the number of declared SEP families for all SDOs is still very significant.[11] Furthermore, as families are defined with respect to common priority documents, whereas essentiality is defined by the relationship between the patent's claims and the standard specifications, one patent family may include both essential and non-essential individual patents. Applying the result of an essentiality assessment of an individual patent to its entire family is thus susceptible of creating an additional error.[12]

Another policy proposal (also included in the report of the SEPs-Expert-Group (2021)) is to only assess potential SEPs selected by patent holders, requiring that only checked SEPs may be asserted against implementers. It is unclear how many declared SEPs patent holders would choose to submit for assessment under such a policy. In contrast to other patents, few declared SEPs are allowed to lapse early (Baron and Delcamp, 2012; Baron and Pohlmann, 2018). There thus does not seem to be that many declared potential SEPs that quickly and unambiguously turn out to be non-essential. Furthermore, many

[10] By comparison, experts assessing essentiality ratios in selected samples in the context of court cases reported to have spent eleven hours per patent preparing claim charts and determining patent essentiality (Cooper (2019), cited from Mallinson (2021)); and Ericsson testified that it spent 50 hours per patent preparing claim charts in TCL v. Ericsson (Mallinson, 2021)

[11] There are different definitions of patent families, which all regroup patents that share one or several priority documents. Using the particulary broad *inpadoc* patent family definition, Baron and Pohlmann (2018) identified 38,800 different patent families in their data (implying a ratio of 3.6 declared SEP per declared SEP family). ETSI, using its own family definition, states that its own database currently includes more than 62,000 declared SEP families. https://ipr.etsi.org/ as of 1 November 2021

[12] If patents subjected to assessments are randomly drawn within each family, this additional error is a random sampling error. If, as suggested in the contribution of the SEPs-Expert-Group (2021), family members are selected using non-random criteria such as the nationality or filing date of the patent, this procedure may also generate systematic bias.

SEPs asserted in court are found to be non-essential, including SEPs asserted by patent holders with very large portfolios of declared SEPs, who surely possess at least some truly essential patents that could have been asserted instead (Lemley and Simcoe, 2018). This fact suggests that patent holders themselves do not know which of their declared SEPs are actually essential. If patent holders themselves are uncertain which of their patents are truly essential, the potential for self-selection into a costly examination process to ”weed out” non-essential patents is limited.

Patent-by-patent essentiality assessments of declared SEPs are thus subject to significant limitations. The cost of very thorough assessments of every declared SEP appears to be prohibitive. While limiting checks to one patent per family and/or only those declared SEPs self-selected by their owners may reduce the number of required assessments, the number of patents to be checked would likely still be significant.[13] Furthermore, the non-random nature of these selections may introduce statistical bias. Reducing the cost of the assessment by reducing the stringency of the assessment would introduce additional measurement error, which may result in further loss of precision and systematic bias.

2.2 Sampling

As an alternative to checking every declared SEP (or one patent for every declared SEP family), it is possible to assess the essentiality of smaller, randomly drawn samples of declared SEPs, and to extrapolate the share of actual SEPs in different firms' portfolios of declared SEPs in the sample to the entire population. This approach was e.g. recommended in studies prepared for the European Commission (Regibeau et al., 2016), and has repeatedly been used in the resolution of licensing disputes through litigation.

Existing applications of sampling approaches in SEP dispute resolution have been subject to significant criticisms.[14] Nevertheless, these criticisms of individual assessments must be distinguished from an analysis of the general merits of sampling approaches. Holding the total resources allocated to an essentiality assessment constant, assessing randomly drawn sample rather than the entire population of interest allows an expert to allocate greater efforts to the analysis of each individual patent. In principle, more thorough assessments can be presumed to be subject to lower measurement error - while even a very thorough assessment may be subject to measurement error, placing resource and time constraints on the assessment process is likely to create an *incremental* measurement error.

[13] Assuming that patent owners would submit the (vast) majority of the declared SEP familiesit is reasonable to expect that more than 50,000 patents would need to be checked, for an assessment of essentiality in the existing stock of declarations alone.

[14] In the TCL v Ericsson case, TCL commissioned subject matter experts to conduct a study of a random sample of 2,600 ETSI declared 2G, 3G and 4G patents. It was calculated that the commissioned experts must have spent on average only about 20 minutes per patent and charged on average $100 per patent for their assessment. The time spent and amount paid for SEP determination for this litigation case very much differed to fees charged and to time spent for essentiality checks e.g. by experts appointed by pools. These resources appear incompatible with a thorough analysis of the relationship between a patent and complex technical specifications that may have up to 600 pages and hundreds of sections. Another critcism of the approach concerned the potential for explicit bias of the experts who conducted the patent mapping. The experts retained by TCL knew which side they were on.

Sampling on the other hand creates a sampling error - the error that results from extrapolating information from a sample to the entire poplulation from which the sample is drawn. Conducting more thorough assessments of a randomly drawn sample rather than a more cursory assessment of every member of a population thus trades measurement error for sampling error. Which of these methods is more *accurate* on average partly depends on the relationship between assessment effort and measurement error, i.e. by how much the accuracy of the assessment of an individual patent increases in effort.

More importantly, the usefulness of light-touch assessments of every individual patent depends on the *distribution* of assessment errors. Measurement errors may lead to systematic biases between different firms' portfolios for at least three reasons: first, experts may give systematic preference to the patents of a certain patent holder (e.g. the firm paying for the assessment); second, Type I errors (a non-essential patent is found to be essential) and Type II errors (an essential patent is found not to be essential) may correlate with different patent characteristics that are non-randomly distributed over firms's portfolios;[15] and third, expert assessments may generally over- or under-estimate the true share of actual SEPs. The percentage of declared SEPs found to be essential in different studies varies widely,[16] suggesting that - depending on the study's methodology and exact criteria of essentiality that are applied - essentiality checks may be overly optimistic or pessimistic about patents' actual essentiality, on average.[17] This may produce biases in the assessment of different firms' portfolios if *uncertainty* regarding patents' essentiality differs between firms.[18]

While at least part of measurement error is likely to be noise, measurement error is thus likely to have a systematic component, which does not vanish when the size of the sample increases. This systematic error (statistical bias) may systematically favor one firm over another. Sampling error, by contrast, is axiomatically unbiased - it does not systematically favor one firm over another. Compared to a cursory assessment of every single patent, an estimate based on the rigorous assessment of a randomly drawn sample, while subject to an additional sampling error, is less subject to this potentially systematic measurement error, and hence less susceptible to bias.

Sampling errors, however, may get very large. While Regibeau et al. (2016) contend that assessments of a sample comprising merely 30 patents may produce adequately precise estimates of the essentiality ratio in the population from which the sample was drawn,

[15] e.g. because experts are overly optimistic or pessimisting regarding the essentiality of patents in different technological fields, and different firms' patents are concentrated in different technological aspects of a standard

[16] See *supra* note 5.

[17] Some assessments are explicitly based on eliminating those declared SEPs that are *clearly not* essential, whereas other assessments aim at counting those patents that positively are essential.

[18] e.g. fundamental patents with broad claims and recent patents with narrow claims that are tailored to the standard may have similar average probabilities of being essential, but it may be easier to ascertain whether the narrow, tailored claim is or is not essential; an overly optimistic review that only discards those patents that are clearly not essential will be overly favorable to a firm holding patents with broader claims whose essentiality is inherently more uncertain, whereas a review that only identifies those patents that clearly are essential and discards all the other ones will be biased in favor of firms with narrower claims that are easier to map to the standard.

Mallinson (2021) argues that "a sample size approaching 3,000 declared-essential patents per standard, at the very least, would be required".[19]

2.3 Predictive modeling

More recently, some authors (Brachtendorf et al., 2020) have proposed a different approach. Other researchers (Stitzing et al., 2017) had already found that the actual essentiality of declared SEPs correlates with more easily observable characteristics of the patent or declaration (e.g. whether the declaration mentions a specific section of the standard for which the patent is alleged to be potentially essential). In addition, Brachtendorf et al. (2020) propose a semantic score measuring the semantic distance between the patent claims and standard specifications, and find that it significantly correlates with the likelihood that a declared SEP is actually essential. Using these and other variables, one can use a sample to estimate an empirical model predicting a patent's essentiality.

Different types of models may be used in this context. While Brachtendorf et al. (2020) estimate a logistic regression, Rangan and Yonamine (2021) propose machine-learning (ML) approaches to essentiality assessments. Similarly to logistic regression, ML uses a sample (or "training dataset") to estimate the relationship between declared SEPs' observable characteristics and an expert's assessment of these patents' actual essentiality, and then makes predictions for essentiality rates in the entire population of declared SEPs.

The specific application of predictive modeling to the assessment of essentiality rates in large populations of declared SEPs is insufficiently understood. Bekkers et al. (2020) e.g. note that the consistency between the predictions of Brachtendorf et al. (2020)'s model and experts' assessment of essentiality is quite low, suggesting that their automated approach is of limited use for making predictions for individual patents. While this is probably true (at least at the moment) for all automated approaches, the relative merits of different approaches for the determination of essentiality rates in larger groups or populations of patents has not been explicitly investigated.

Similar to sampling, predictive models can generally be expected to improve with the number of observations in the sample. Nevertheless, different estimation methods have different characteristics. Pure sampling of patents uses only the patents of a specific firm in a sample to predict the essentiality rate in that firms' patent portfolio. Methods using explanatory variables may draw from the entire sample (not limited to the patents of the specific firm) to predict the essentiality rate in each individual firm's portfolio. Therefore, these methods have the potential to yield more precise predictions, in particular for essentiality shares in small portfolios. Nevertheless, including or excluding a specific variable may systematically advantage or disadvantage individual firms.

The performance and statistical properties of predictive modeling approaches obviously depend on the specifics of each approach, i.e. the number and identity of explanatory

[19]While the authors also disagree about the correct application of basic statistical theory, the authors apply different assumptions that contribute to explain these discrepancies: Mallinson (2021) incorporates assessment inconsistencies into the calculation of the confidence intervals, assumes a low essentiality rate among declared SEPs, and defines an acceptable size of the confidence interval as a fraction of the estimated essentiality rate (for an estimated essentiality rate of 10%, an acceptable confidence interval under his definition would span from 8.5 to 11.5%, a mere three percentage points interval.)

variables, the type of model (e.g. logistic regression or tree-based ML algorithms), the functional form of the regression equation, etc. In our analysis, we do not aim to exhaustively assess all plausible predictive modeling approaches. Nevertheless, we aim to achieve several objectives: first, we broadly replicate the fundamental tenets of the analysis of Brachtendorf et al. (2020), the (so far) only predictive modeling approach that is discussed in a publicly available paper in sufficient detail, in a different sample, for a more recent technology standard generation, and using essentiality assessments by different experts. Second, we aim to improve upon this approach using our more comprehensive database of promising explanatory variables, and a more precise assessment of the essentiality of declared SEPs for specific TS of a complex standard. Third, we aim to assess the out-of-sample performance of predictions of essentiality rates in different firms' portfolios, and to derive some generalizable take-aways regarding the promise, limitations, and best use of predictive modeling approaches to SEP determinations more generally.

3 Data

For the empirical analysis, we use declarations of (potential) SEPs from the ETSI IPR Database. The analysis considers all patent declarations published at ETSI up until October 1, 2020 and classified as 5G relevant. Patent declarations were classified as 5G relevant if the Technical Specifications (TS) of the declaration were marked as 5G technology by the 3GPP. For the analysis, we only take into account patents that were granted at the USPTO (United States Patent and Trademark Office) or granted at the EPO (European Patent Office) by October 1, 2020, and that were active (not expired, revoked or lapsed by October 1, 2020). Furthermore, we restrict the analysis to patents currently assigned to one of 35 firms (or their fully owned subsidiaries), which are among the largest holders of declared SEPs complying with our different sampling criteria. These 35 assignees account for 19,488 active EP and US patents declared as potentially essential to a 5G TS by October 1, 2020 (more than 95% of the total number of such patents in the ETSI database).

For the purposes of sampling, patents were *collapsed* by extended INPADOC family ID; i.e. for each extended INPADOC family, at most one representative patent can be selected into the sample. There are 10,860 different extended INPADOC families in our data, from which we selected a random sample of 960 families.[20] For each INPADOC family, if applicable, the representative patent is randomly chosen between the earliest EP and the earliest US patent (by application date). For INPADOC families only containing US or only containing EP patents, respectively, the earliest patent is the representative

[20]The initial sample included 1,000 patents. We eliminated a small number of INPADOC families with an unusually large number of family members in the sample (more than ten sample patents) from both the sample and the reference population. There are also a small number of cases in which various patents in the sample are now listed as member of the same INPADOC family (presumably because an application filed since creation of the original sample constitutes a common link between two sample patents). While these adjustments (and the drop of 4% of the original sample) may result in the sample not being 100% randomly selected, we took care to apply all adjustements equally to the sample and the reference population. Furthermore, our comparative analysis of different methods is based on truly random draws of subsamples from our sample.

patent. For INPADOC families not selected into the sample, we only use the observable characteristics of the representative patent to infer predicted essentiality rates.

The selected sample patents were examined by our technical experts, who provided an assessment of the likelihood that the patent is essential to 5G standards.[21] Experts classified patents as "fully mapped" (12.4%), "partially mapped", "edge case", or "not mappable" (56.4%).[22] For our analysis, we conflate the two intermediary categories (partially mapped and edge case; together accounting for 31.2%). The rate at which patents are found to be "essential" is thus highly dependent on the specific criterion, ranging from 12.4% patents fully mapped to 43.6% patents fully or partially mapped. In addition to providing an overall assessment of the patent's essentiality to 5G standards in general, for patents partially or fully mapped, experts provided the identifier of the relevant technical specification (TS) to which the patent was mapped. We matched this patent-TS mapping to the declaration data. Disregarding different versions of the same TS, patents in our sample were declared essential to - on average - 3.28 different TS. The experts could only confirm 17.8% of these declared potential essentiality relationships; on average, for each patent there are 0.58 TS for which the patent was both declared and found to be essential. In addition, the experts mapped patents to 0.197 TS, on average, for which they had not been declared.

4 A regression framework to predict essentiality

In this section, we will estimate each sample patent's probability of being essential as a function of the patent's observable characteristics. To this end, we perform a set of logit regressions in the sample to estimate the relationship between patents' observable variables and mapping status; and apply the regression coefficients to the entire patent population to calcucate predicted essentiality probabilities.

As explained above, a patent is essential to 5G if it is essential to at least one 5G TS. In most but not all cases in which a patent is found to be essential to a TS, it has been declared to be potentially essential to this specific TS.[23] As a first step, we thus analyze for all declared SEPs in the sample whether they are effectively essential to the specific TS to which they were declared to be potentially essential.

[21] Essentiality assessments were ordered and paid for by iplytics, but independently carried out by unaffiliated subject matter experts.

[22] These categories were defined by our technical experts, who describe the categories as follows:

- "Fully Mapped: All the claim elements were found in standards, chart made to justify that the patent is relevant (100% Mapping).
- "Edge Case": All the claim elements found in the standards (95%) except a very specific element not explicitly disclosed in the standard. Left to interpretation. It may become full mapped or may remain partial.
- "Partially Mapped": Most of the claim elements were found in standards, except one or two concepts, chart made to justify that the patent is relevant (More than 60% Mapping).
- "Not Mappable": All the claim elements were not found in standards and patent is found not relevant (If less than 50% Mapped).

[23] There may also be patents that are actually essential to a specific 5G TS, but have not been declared potentially essentialy to *any* 5G TS. These patents are completely excluded from our analysis.

For this analysis, we can use a substantially larger sample of more than 3,000 patent-TS-relationships, as each patent was declared to be essential to - on average - more than three different TS. In addition, we can make use of highly granular patent-TS-relationship specific variables, which are highly plausible candidate variables for predicting effective essentiality.

For some of these variables, we can directly build on the existing literature. Both Stitzing et al. (2017) and Brachtendorf et al. (2020) find that patents that were declared to be potentially essential to a specific section of a standard specification have a significantly higher likelihood of actually being essential. Presumably, such section-specific declarations are based on more specific, and hence more reliable, information and beliefs. Brachtendorf et al. (2020) find that patents whose claims have a higher semantic similarity with the text of a standard specification have a higher likelihood of being standard-essential.

We therefore use our declaration data to identify patent-TS-declarations referencing specific TS sections. About 7.2% of the declarations in our data are section-specific (i.e. among all patent-TS-pairs in our data, for 7.2% of the pairs, there is at least one section-specific declaration; see Table 2). Furthermore, we calculate similarity scores between the independent claims of all declared SEPs and the TS to which they were declared essential.[24] These variables are not only a priori plausible and supported by the existing literature, they also exhibit a strongly significant positive correlation with the mapping status of the patent - the likelihood that the experts found a declared SEP to be effectively essential to a TS increases when the patent was declared essential to a specific section of that TS, and when there is a higher similarity score between the language of the patent's independent claims and the text of that TS (Table 4).

Variable	**Mean**	**Std. Dev.**	**N**
standard_cited_npl	0.281	0.449	2818
prioritydateoverlap	0.089	0.057	3174
related_wg_attendance	35.721	57.665	3106
related_contributions	2.417	20.67	3174
cpcoverlap	0.034	0.039	3174
semantic_score	0.66	0.05	3069
sectiondeclared	0.072	0.258	3183

Table 2: Summary statistics: Patent-TS relationship characteristics

[24] We use a Latent Semantic Indexing (LSI) model to calculate similarity scores. While this analysis is not a direct replication of the proprietary semantic analysis used by Brachtendorf et al. (2020), our results appear to be similar; and also Brachtendorf et al. (2020) find that their preferred algorithm produces results that are similar to those of other, freely available semantic analysis methods. While innovations in semantic analysis may improve the ability of semantic indicators to predict an expert's finding that a patent is indeed essential, we expect that our broader findings are generalizable to other methods based on comparing the text of patent claims and standard specifications to calculate a metric of similarity. See Appendix for details.

Variables	mapped	npl	date	attend	contr	cpc	semantic	section
mapped	1.000							
standard_cited_npl [npl]	0.098	1.000						
	(0.000)							
prioritydateoverlap [date]	-0.090	-0.090	1.000					
	(0.000)	(0.000)						
related_wg_attendance [attend]	0.184	0.087	-0.059	1.000				
	(0.000)	(0.000)	(0.001)					
related_contributions [contr]	0.105	0.038	-0.010	0.153	1.000			
	(0.000)	(0.041)	(0.557)	(0.000)				
cpcoverlap [cpc]	0.062	-0.030	0.512	0.123	-0.012	1.000		
	(0.000)	(0.114)	(0.000)	(0.000)	(0.493)			
semantic_score [semantic]	0.191	0.053	-0.091	0.157	0.087	0.087	1.000	
	(0.000)	(0.006)	(0.000)	(0.000)	(0.000)	(0.000)		
sectiondeclared [section]	0.126	0.087	-0.066	0.044	0.032	-0.017	0.041	1.000
	(0.000)	(0.000)	(0.000)	(0.015)	(0.069)	(0.351)	(0.022)	

Table 4: Pairwise correlation table: Patent-TS relationship characteristics

The similarity of our findings to those of Stitzing et al. (2017) and Brachtendorf et al. (2020) provides some reason to be optimistic about the prospects of predicting the essentiality of declared SEPs using observable characteristics. The fact that the same or closely related variables consistently predict expert findings of essentiality in different samples (drawn from different standard generations) suggests that the relationship between these patent characteristics and the patents' mapping status can indeed be generalized to larger populations of declared SEPs. The fact that these variables consistently correlate with different experts' assessment of essentiality furthermore suggest that the correlation extends beyond one specific method of evaluation, and may be generalizable to other assessments of comparable scope and depth.

We complement these knwon predictors of essentiality with another set of variables. While Brachtendorf et al. (2020) find that the number of prior art citations from a patent to non-patent literature predicts essentiality for 4G (LTE), but not for 2G (GSM) or 3G (UMTS) patents, we test for the more specific patent-TS-relationship specific variable indicating whether the patent cites the specific TS to which it was declared to be essential as part of the NPL citations. We find this to be the case for more than 28% of the patent-TS pairs in the sample, and the variable correlates very significantly with the patent being found to be essential to this specific TS.

In addition, we measure the *CPC overlap* between the patent, and all other patents declared potentially essential to the same TS. Patents sharing a higher number of CPC classifications with a higher proportion of the population of declared SEPs for this particular TS are more likely to be technologically closely related to this TS - as confirmed by a statistically highly significant positive correlation with the likelihood of being found essential. We similarly calculate the priority date overlap between this patent and all other patents declared potentially essential to this TS. We find that patents with a priority date that is close in time to the priority date of many other patents declared potentially essential to the same TS are statistically significantly *less* likely to be found essential to this TS - perhaps an indication of greater technological uncertainty at times of significant inventive activity; or peaks in strategic or opportunistic patenting and declaration of "just-in-time" patents (Kang and Bekkers, 2015).

We furthermore build patent-TS-relationship specific variables from our extensive databases with 3GPP contribution and attendance information. In particular, *related_wg_attendance* measures attendance by the patent inventor(s) in the working group(s) related to the technical specification to which the patent was declared potentially essential. Kang and Motohashi (2015) find that participation of the inventor in 3GPP meetings is a significant predictor of the patent being declared to be *potentially* essential. Our results demonstrate that 3GPP meeting attendance by the inventor furthermore is significantly correlated with the likelihood that a patent is actually found to be essential, conditional on it being declared as potential SEP. Similarly, we find that a count of *related_contributions* significantly correlates with the likelihood that a declared SEP is actually found to be essential. These findings may suggest that firms (and inventors) that directly and intensively participate in 3GPP have a better understanding of the standard, and thus are able to make more accurate declarations of potential essentiality. The findings do not imply however that contribution and/or attendance counts are direct measures of substantive technical contributions to standards development.[25]

We combine these different independent variables into a predictive model, predicting the probability that a patent is actually found to be essential to a specific TS for which it has been declared to be potentially essential (Column (1), Table 5).[26] All variables are individually statistically significant, suggesting that they each add information that is useful for predicting essentiality of a patent to a TS.

[25]See **?** for a critical assessment of the use of contribution counts in the assessment of SEP portfolio value.

[26]In most of our analyses, we consider that all patents "mapped" or "partially mapped" have been found "essential". We explore some of the implications of the distinction between "fully mapped" and "partially mapped" in Section X below.

	(1)	(2)	(3)	(4)
	mapped	atleastonemapped	atleastonemapped	atleastonemapped
	Patent-TS	**M1**	**M2**	**M3**
standard_cited_npl	0.468***		0.448**	0.360*
	(4.07)		(2.78)	(2.04)
prioritydateoverlap	-6.276***		-5.265*	-4.376*
	(-4.38)		(-2.47)	(-2.13)
related_wg_attendance	0.00253***		0.00362**	0.00226
	(3.39)		(3.04)	(1.80)
related_contributions	0.00447		0.00258	0.00383
	(1.91)		(1.12)	(1.48)
cpcoverlap	8.350***		6.077*	4.730
	(5.40)		(2.42)	(1.96)
semantic_score	5.936***		5.995***	5.303**
	(5.38)		(3.91)	(3.23)
sectiondeclared	0.442**		0.447	0.215
	(2.62)		(1.92)	(0.59)
numberspecsdeclared	-0.211***	0.0657*	-0.0346	-0.00122
	(-8.93)	(2.40)	(-1.35)	(-0.04)
total_phat		3.862***		
		(7.08)		
technicalrelevance		0.0831*	0.0794*	0.0772*
		(2.22)	(2.17)	(2.15)
radicalness		0.0371	0.0475	0.0324
		(1.43)	(1.72)	(1.10)
Constant	-4.573***	-1.912***	-4.689***	-4.366***
	(-5.90)	(-7.53)	(-4.34)	(-3.56)
Observations	2,642	835	835	813
Firm FE	N	N	N	Y

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Logit regression - Probability of patent being essential

There are many different measures for the performance of predictive models. As a first step, we plot the mapping status of patent-TS-pairs by predicted probability brackets - i.e. we divide the sample of 2,642 patent-TS-pairs that was used in the regression into 20 brackets of equal size, ranked from the top-5-percentile to the bottom-5-percentile in terms of predicted essentiality probability; and then count the number of actually mapped pairs within each bracket. The results in Figure 1 indicate that the model is highly accurate in ruling out essentiality for a relatively small share of the population of patent-TS-pairs - the bottom 20% of the distribution indeed only include very few observations of patents that were actually found to be essential to this particular TS. The converse, however, is not true - only in the very first Top 5% bracket, more than half of the declared patent-TS-pairs were actually confirmed by the technical experts, and only by slight margins. The predictive model thus does not identify any sample of patent-TS-pairs for which essentiality can be presumed with high degreees of confidence.

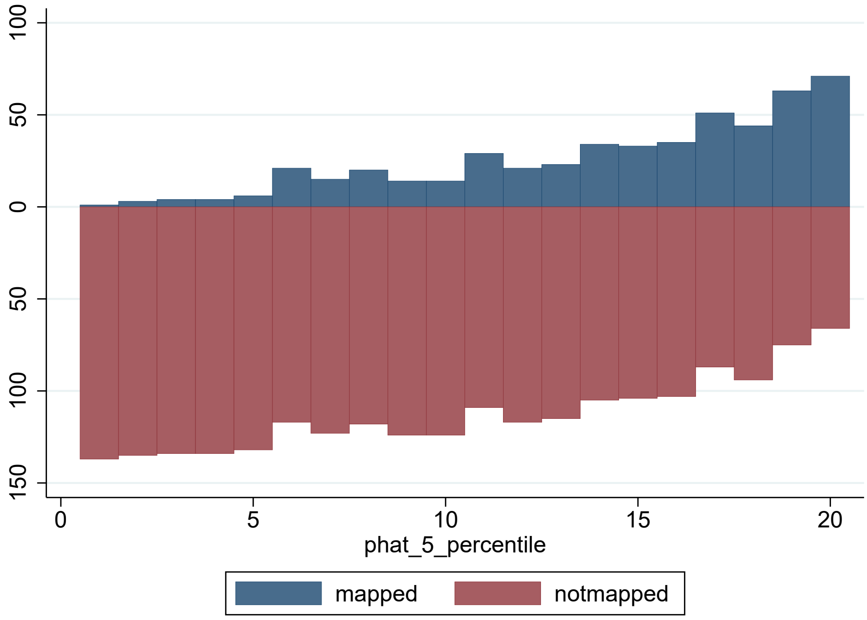

Figure 1: Mapping status by predicted probabilities; patent essential to TS for which it was declared

While the results in Figure 1 present the entire continuum of predicted probabilities, many standard indicators for the performance of predictive models are based on binary predictions. In Table 6, we display the *confusion matrix* for a cutoff at a probability of 0.5 (i.e. a predicted probability of 0.5 or higher indicates a ”positive” prediction). The table plots the numbers of false and true positive and negative predictions (e.g. false positive predictions are those patent-TS-pairs for which the predicted probability based on the predictive model is >0.5, but the expert’s assessment is ”unmapped”).

	unmapped	mapped	all
prob<0.5	2131	441	2572
prob>=0.5	27	43	70
all	2158	484	2642

Table 6: Confusion matrix: patent essential to TS for which it was declared

So far, we have analyzed patent-TS-pair specific variables, and their usefulness for predicting whether a patent is essential to a specific TS to which it was declared. In many settings, we will be more interested in predicting whether a declared SEP is essential to *any* 5G TS. We can make use of some patent-level information, i.e. variables that vary from one patent to the other, but do not change between different TS. Several of these variables are based on patent citations. Counts of forward citations, i.e. the number of prior art citations that a patent receives from ulterior patents, are a commonly used indicator of patent quality or significance (Trajtenberg, 1990; Hall et al., 2005), and it has repeatedly been shown that patents declared to be *potential* SEPs receive significantly higher numbers of forward citations than other, otherwise comparable patents (Rysman and Simcoe, 2008; Bekkers et al., 2017; Baron and Pohlmann, 2018). Rysman and Simcoe (2008); Bekkers et al. (2017) furthermore find that rates of forward citations significantly increase after a patent’s declaration to an SDO, suggesting the existence of a ”disclosure effect". Brachtendorf et al. (2020) find that the number of forward citations significantly predicts that a patent declared to be potentially essential is actually found to be essential, and that the "disclosure effect" discovered by Rysman and Simcoe (2008); Bekkers et al. (2017) depends on the patent's semantic similarity to the standard, a measure of actual essentiality. Taken together, these findings suggest that forward citations are correlated with a patent's actual essentiality.[27]

The count of patent citations is subject to widely studied methodological considerations. Counting citations from individual patents may result in over-counting, as different patents relating to the same invention are likely to cite the same prior art. Furthermore, citation counts may be biased towards older patents, as these patents had more time to accrue citations. Finally, citation counts may include significant numbers of self-citations (i.e. citations from other patents of the same owner), which are potentially subject to strategic behavior. To address these considerations, we use a *technical relevance* indicator, an age-normalized count of forward citations received from different independent patent families, excluding self-citations.

Another citations-based indicator we use is *radicalness*; an inverse measure of the number of backward citations, i.e. the citations from the focal patent to other patents. Backward citations have sometimes been found to be *positively* correlated with patent value (Lanjouw and Schankerman, 2004); nevertheless, our measure of radicalness may indicate that a patent protects an invention that is more distinct from the prior art. Empirically, we find that both *technical relevance* and *radicalness* are positively correlated with a positive essentiality assessment (i.e. the patent is found to be essential to at least one 5G TS), see Table 8.

Furthermore, we use an indicator of *legal breadth*, which is the number of words in the shortest independent claim of the patent, normalized by jurisdiction and CPC class. The length of the shortest independent claim is an increasingly accepted measure of claim "scope" or "breadth" of the legal protection awarded by a patent (Marco et al., 2019). On average, the addition of more words to a patent claim makes the claim more specific, and hence narrower. Brachtendorf et al. (2020) find that the length of the first claim of the patent is negatively correlated with the likelihood that a patent declared to be potentially essential is found to be actually essential by an independent expert. In our data, the length of the shortest claim is negatively correlated with an essentiality finding, but the correlation is only mildly significant (significant at 10% but not at 5%). Other relatively commonly used patent characteristics, such as family size (the number of patents in the INPADOC family), and team size (the number of inventors listed on the patent) are not correlated with our experts' assessment of declared patents' actual essentiality.

[27]There can be different mechanisms which can potentially explain this correlation - follow-on inventors may be more likely to "build on" inventions incorporated into standards; and patents receiving many forward citations may be intrinsically more fundamental and difficult to circumvent (characteristics that also increase the probability of the patent being essential).

Variables	atleastonemapped	specs	tr	rad	cits	LB	team)	family
atleastonemapped	1.000							
numberspecsdeclared [specs]	-0.043	1.000						
	(0.178)							
technicalrelevance [tr]	0.090	-0.001	1.000					
	(0.005)	(0.973)						
radicalnessra [rad]	0.077	0.005	0.004	1.000				
	(0.018)	(0.866)	(0.913)					
Forward Citations [cits]	0.026	-0.037	0.360	-0.042	1.000			
	(0.413)	(0.244)	(0.000)	(0.199)				
Legal Breadth [LB])	-0.060	-0.001	0.111	0.040	0.043	1.000		
	(0.061)	(0.985)	(0.001)	(0.213)	(0.179)			
Team Size [team]	0.031	-0.054	0.025	-0.041	0.029	0.034	1.000	
	(0.342)	(0.091)	(0.434)	(0.211)	(0.370)	(0.294)		
Family Size [family]	0.009	-0.021	0.010	-0.224	0.061	0.071	0.053	1.000
	(0.785)	(0.513)	(0.761)	(0.000)	(0.057)	(0.027)	(0.102)	

Table 8: Pairwise correlation table: Patent characteristics

In addition to these patent-level characteristics, we hypothesize that the likelihood that a patent is essential to any TS depends on the characteristics of its relationship with each of the TS to which it was declared to be potentially essential. Specifically, a patent is essential to 5G if it is essential to at least one 5G TS. We can thus express the probability Π_i that patent i is essential to 5G as a function of the probability $\pi_{i,k}$ that patent i is essential to any of the N specifications $k = 1, 2, ..., n$ to which the patent is *potentially* essential:

$$\Pi_i = 1 - \prod_{k=1}^{n}(1 - \pi_{i,k}) \tag{1}$$

Empirically, we can estimate $\hat{\pi}_{i,k}$ using the regression equation $\pi_{i,k} = \alpha_1 X_{i,k} + \epsilon_{i,k}$, where $X_{i,k}$ is the vector of our patent-TS-pair specific variables. Using the $\hat{\pi}_{i,k}$ thus estimated, we can calculate $\hat{\Pi}_i$ using equation (1). $\hat{\Pi}_i$ is a valid estimation of Π_i if error terms $\epsilon_{i,k}$ are uncorrelated across the different specifications for which a patent has been declared. We however expect that there may be unobserved variables simultaneously affecting the probability that a patent is essential to various specifications. Furthermore, a patent may be found to be essential to a specification to which it was not specifically declared. To allow for these different considerations, we estimate a second regression equation:

$$\Pi_i = \alpha_1 \hat{\Pi}_i + \alpha_2 Y_i + \epsilon_i \tag{2}$$

where Y_i is a vector of patent-specific characteristics. We can think of the estimated $\hat{\Pi}_i$ as a linear transformation of different patent-TS-specific characteristics into a single patent-specific variable. We will refer to this specification as ”Model 1”, or M1. Alternatively, and in line with Brachtendorf et al. (2020), we will estimate Π_i directly as a function of patent-specific characteristics, as well as characteristics specific to the relationship between patent i and TS $\tilde{k}$, which is the TS most similar (in terms of semantic similarity score) to the patent (among the TS to which the patent was declared).[28] We will refer to this alternative specification as "Model 2", or M2:

$$\Pi_i = \alpha_1 Y_i + \alpha_2 X_{i,\tilde{k}} + \epsilon_i \tag{3}$$

Finally, we test an additional predictive model, which is equivalent to M2, except that it adds firm fixed effects to the regression equation. One can think of a model with firm fixed effects as a hybrid between pure sampling and predictive modeling. While pure sampling exclusively uses a firm's essentiality ratio in the sample to predict the same firm's out-of-sample essentiality ratios, predictive modeling uses observable patent characteristics, and the relationship between these observable characteristics and patents' probability of being essential in the sample, to predict the essentiality of patents not included in the sample. M3 relies on both observable patent characteristics and firms' in-sample essentiality rate to make out-of-sample predictions.

The results of the estimations of Models 1, 2, and 3 are presented in columns (2), (3), and (4), respectively, of Table 5. In Table 9, we present the confusion matrices associated with Models 1, 2, and 3. In Figure 2, we display the number of patents that were and that were not mapped to at least one 5G TS over the distribution of predicted probabilities, comparing Models 1, 2, and 3. We can see that the performance of the different models is very similar. Model 3, which uses the largest number of covariates, makes the most accurate in-sample predictions - in particular, it stands out as the Model that is most capable of identifying a (small) group of patents with a very high likelihood of being found essential.

	unmapped	mapped	all
M1.prob<0.5	391	222	613
M1.prob>=0.5	87	135	222
M2.prob<0.5	389	206	595
M2.prob>=0.5	89	151	240
M3.prob<0.5	373	191	564
M3.prob>=0.5	95	154	249
all	478	357	835

Table 9: Confusion matrix: patent essential to at least one 5G TS

We display a summary of prediction performance measures in Table 10, such as sensititivity (or "true positive rate" or "recall", i.e. the share of positive observations correctly predicted), specificity (or "true negative rate", i.e the share of negative observations correctly predicted), positive predictive value (or "precision", or the share of positive predictions that are true positives) and negative predictive value (the share of negative predictions that are true negatives). All these measures are affected by the cutoff level and the distribution of outcomes.

[28] More specifically, we first eliminate patent-TS-observations with missing information for one of the variables used in the regression; $\tilde{k}$, is thus the TS most similar to the patent among the TS to which the patent was declared, and for which we have all information required to compute predicted essentiality.

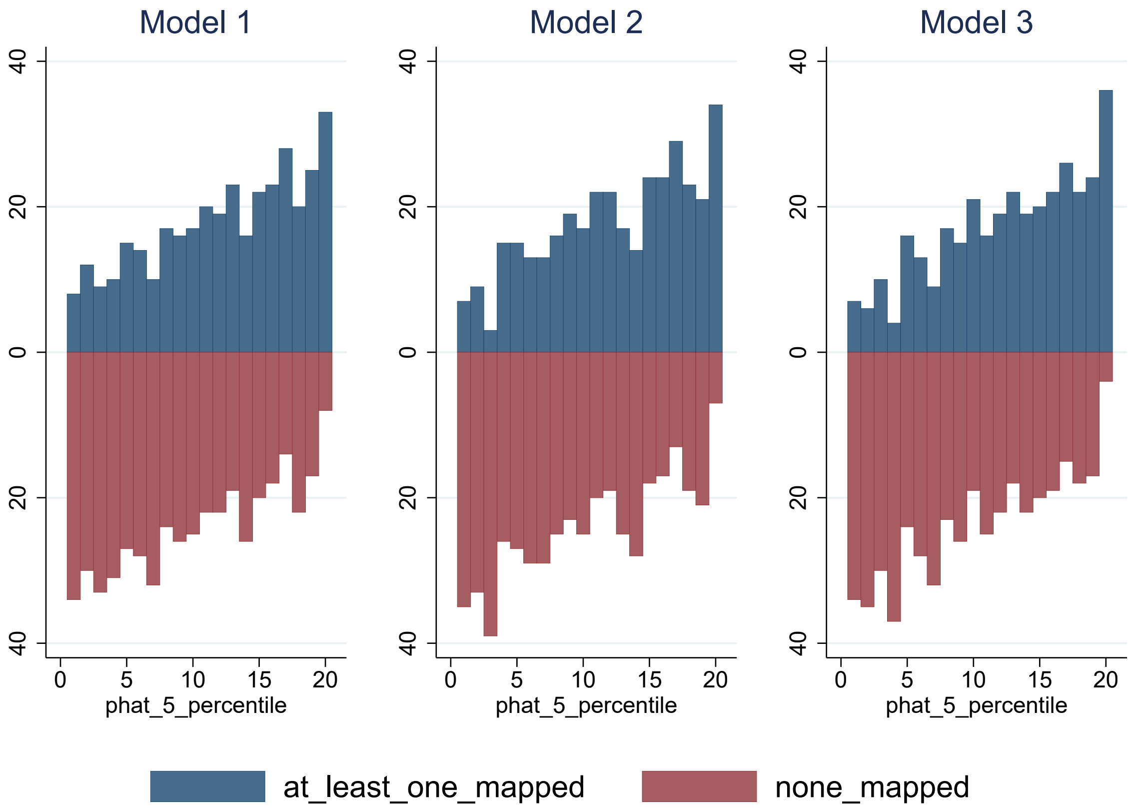

Figure 2: Mapping status of patents (mapped to at least one 5G TS), by predicted probability of essentiality

A more informative measure of overall prediction performance is Matthews' correlation coefficient (MCC), which is independent of the distribution of observations between positive and negative outcomes. A MCC equal to 1 indicates that the model predicts each outcome perfectly, whereas a MCC of 0 indicates that the model is equivalent to a random prediction. The MCC once again confirms that Model 3 makes the most accurate predictions, and that our Model 1 is significantly dominated by the more direct patent-level estimations of essentiality - at least in-sample.

Nevertheless, these performance measures only represent intermediate assessments - we are not seeking to predict essentiality rates in the sample, but in the un-tested population of patents; and we are less interested in predicting the essentiality probabilities of individual patents than essentiality rates in different firms' patent portfolios. A different set of measures is required to assess the usefulness of different models in this context.

	Patent-TS-pair	Patent		
		M1	**M2**	**M3**
pct correctly classified	82.25	62.99	64.67	64.82
sensitivity	9.917	37.82	42.3	44.64
specificity	98.47	81.8	81.38	79.7
pos predictive value	59.26	60.81	62.92	61.85
neg predictive value	82.98	63.78	65.38	66.13
Matthews corr coeff.	.1882	.2196	.2588	.261

Table 10: Prediction performance statistics

5 Comparison of predictive modeling and sampling

5.1 Comparison of different predictions for SEP population

In this section, we will use our sample of checked declared SEPs to make predictions about essentiality rates in firms' overall portfolios of declared 5G patents. In particular, we will compare predictions based on sampling methods and predictive modeling.

There is a total of 10,874 INPADOC families that comply with all sample criteria.; 989 INPADOC families out of this population were randomly selected for an essentiality assessment. The remaining (unchecked) comparison sample thus includes 9,885 different INPADOC families.[29]

When using our regression results to predict essentiality rates in the population, we reproduce the sampling criteria that we used for selecting patents for essentiality assessments. We thus identify the *representative* patents within each family. In line with our sampling criteria, these are the earliest EP and US applications within each INPADOC family. Among various family members issued by the same patent office with the same application date, we select the patent(s) with the earliest publication date. For INPADOC

[29]Some INPADOC families may be assigned to multiple firms within our sample of 35 sample firms (either because different family members are assigned to different firms, or because a single patent has multiple assignees, or because an assignee is a joint venture jointly owned by two sample firms). The comparison sample thus consists in 9,945 INPADOC family observations.

families with multiple representative patents,[30] we average the predicted probabilities of all representative patents to calculate the family's predicted probability of being essential. The comparison sample contains a total of 17,796 patents, of which 13,949 are representative patents.

In order to allow for comparisons between different methods, an additional sampling restriction needs to be applied: while we can use 989 INPADOC families for a sampling approach, we can only use the regression approach to predict essentiality rates for those patents for which all explanatory variables are observable. Out of the 10,874 INPADOC families in the population, we have complete information to predict essentiality rates of 9,298 families (including 828 checked sample patent families and 8,470 unchecked families in the comparison sample).

For both approaches, we compute confidence intervals for our predictions. In the sampling approach, each firm' sample patents are a random draw from the respective firm's population patents; we can thus calculate the confidence intervals for the predicted population essentiality rates using the observable sample standard deviation in the sample. Given that - on average - approx. 10% of a firm's patents are selected into the sample and checked, population confidence intervals slightly overstate the extent of uncertainty - there is uncertainty only regarding the share of essential patents in the unchecked part of the firm's portfolio. We can adjust for this using finite population correction factor to calculate confidence intervals for the share of essential patents in firm portfolios (comprising both checked and unchecked patents).

We numerically approximate confidence intervals for the predicted number of essential patents owned by different firms.[31] In a first step, we draw 200 different samples with replacement (bootstrap samples) from the sample of tested patents, where each boostrap sample has the same size as the sample itself. We estimate the two models in each of the bootstrap samples to obtain an approximation of the full distribution of predictions for the essentiality probability of each unchecked patent. In a second step, for each of the 200 predictions, we randomly generate 200 realizations of essentiality by patent, given each patent's individual predicted probability of being essential. From these 40,000 iterations, we observe each firm's probability mass function of the probability that its portfolio contains X essential patents, for each $X \epsilon [0, N]$, which we can use to identify the 90%

[30]Many INPADOC families have representative EP and US members. There can also be multiple representative members from the same office, if these patents have the same application and publication dates.

[31]To compute confidence intervals for essentiality shares predicted using regression approaches is non-trivial. The models predict (with uncertainty) the probability for each individual patent to be essential. Computing the confidence interval for these predictions is different however from computing confidence intervals for the predicted number of essential patents in a firm's portfolio. The probability of each individual patent to be essential defines the probability mass function of the probability that a firm's portfolio contains exactly X essential patents for each X from 0 to N, where N is the number of potentially essential patents declared by the firm. Given that each patent's probability of being essential is different and individually determined, one would need to calculate the probabilities for each of the 2^N possible combinations of essential/non-essential realizations in a portfolio of N patents, and sum up the probabilities of all combinations resulting in the same number of essential and non-essential patents. Given that even for a modest portfolio size of 30 patents, there are over 1 million combinations to calculate, this is not practically feasible.

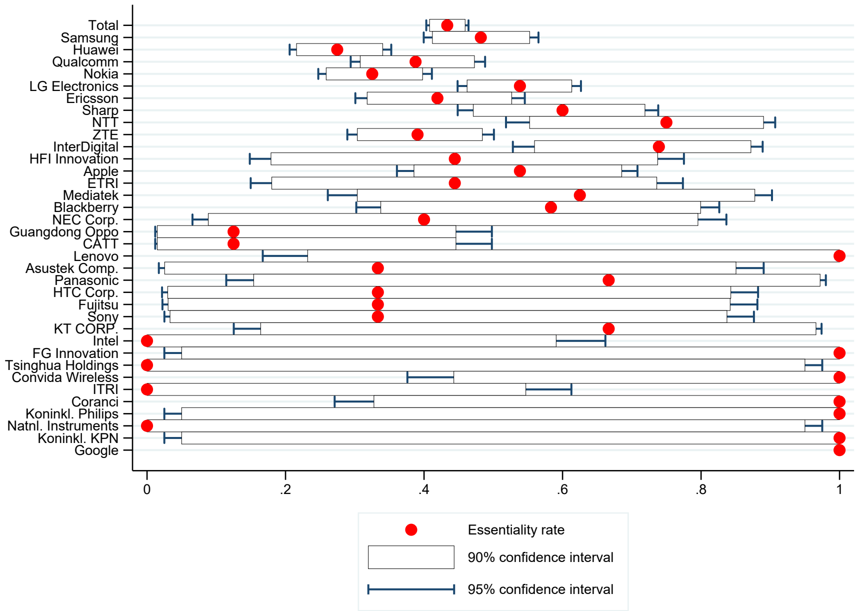

Figure 3: Essentiality rates and confidence intervals (corrected for finite population size) - sampling method

and 95% confidence intervals for the predicted number of essential patents in each firm's portfolio.

We depict predicted essentiality rates based on sampling, along with 90 and 95% confidence intervals, in Figure 3. Sampling provides a fairly precise estimate of the share of essential patents in the total population of declared SEP. With 95% confidence, this share in the total population is contained in an interval ranging from 40.3 to 46.4% - a sizeable, but still reasonable confidence interval. The confidence interval at 90% confidence is slightly smaller (40.8 to 45.9%), whereas the 95% confidence interval based on the smaller subsample of 822 patent families with complete information on covariates is slightly larger (39.7% to 46.3%), but there are no indications that the share of essential patents significantly differs between patents with complete and incomplete information on relevant covariates (this is reassuring, as it suggests that using a regression framework to predict essentiality rates using covariates with missing observations does not inherently introduce an additional sampling bias). The finite population correction factor has a modest effect on the size of the confidence intervals for the total population.

In contrast to the relatively precise prediction of the share of essential patents in the total population, predicted shares of essential patents in individual firms' portfolios are based on much smaller samples, and thus necessarily much less precise. Only for some firms with more than 100 checked patents in the sample, 95% confidence intervals are smaller than 20 percentage points; for firms with small numbers of patents in the sample, the only prediction that the sample allows to make with reasonable confidence is often that their share of essential patents ranges anywhere from single digit percentages to more than 80%. Only for a small number of firms' portfolios, we can conclude with reasonable levels of confidence that the share of essential patents in their portfolio of declared SEPs is signficantly lower (e.g. Huawei, Nokia, ZTE, Guangdong Oppo, CATT) or higher (LG Electronics, Sharp, NTT, InterDigital) than the population average. The finite population

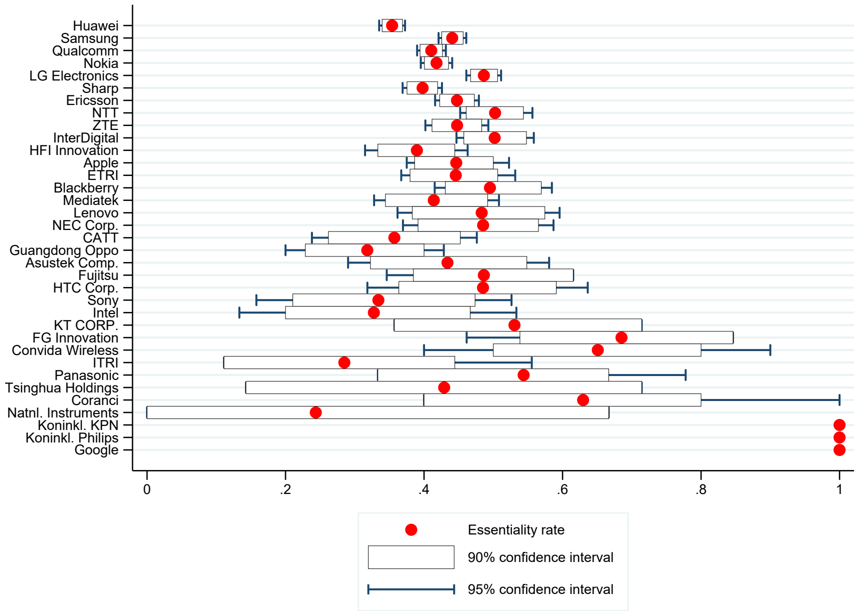

Figure 4: Essentiality rates and confidence intervals - logistic regression method

correction factor significantly reduces the size of the confidence intervals only for some firms, whose share in the sample is disproportionately large compared to their share in the population (e.g. ZTE).

While sample essentiality ratios thus provide relatively precise predictions of population essentiality ratios for the entire population of declared SEPs, a much larger sample of checked patents would be required to produce useful predictions of the essentiality ratios in individual firm portfolios, particularly portfolios of firms with only small numbers of declared patent families. Alternatively, we can use our observable variables to make predictions for essentiality rates in individual firm portfolios. Figure 4 depicts essentiality rates by firm portfolio, based on Model 2.[32]

As can be seen from comparing Figures 3 and 4, confidence intervals for predicted essentiality rates - especially for medium and small portfolios - are much smaller in the predictive modeling approach. This clearly demonstrates that predictive modeling has the potential to produce significantly more precise estimates of essentiality rates in individual portfolios. This advantage is particularly pronounced for smaller portfolios - unlike sampling, predictive modeling uses the entire sample of checked patents, and not only the patents in an individual firm portfolio, to estimate the relationship between essentiality and observable covariates. It is thus capable of making relatively precise predictions for essentiality rates in individual portfolios, even if no or only very few patents from that specific portfolio were checked.

[32] As explained above, we use bootstrapping and random realizations of individual patents' probabilities to be essential to calculate essentiality rates and their confidence intervals. Figure 4 thus depicts the distribution of the predicted shares of firms' patents that are essential, rather than the (much easier to calculate) distribution of individual patents' probability to be essential. We report results based on Model 2 because of its superior *out-of-sample* performance (see section 5.2. below).

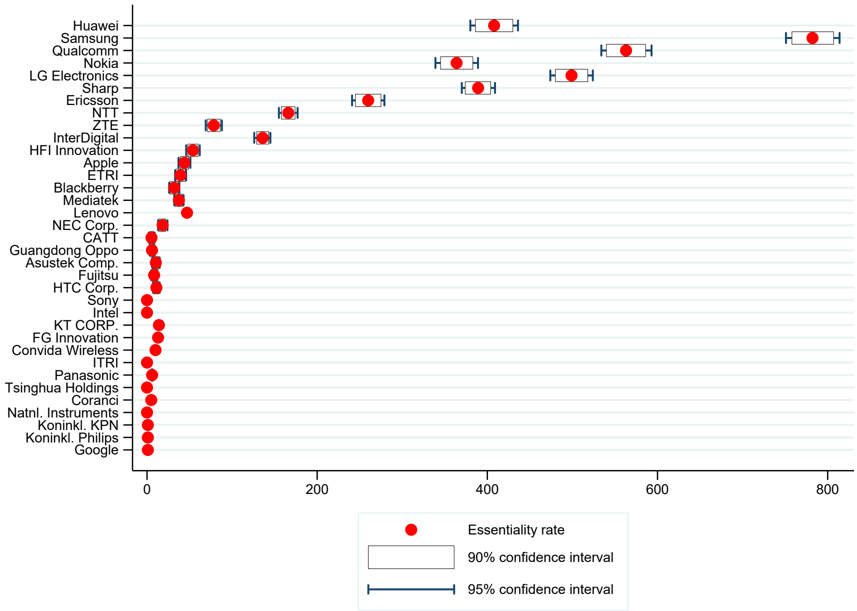

Figure 5: Number of actual SEPs and confidence intervals (corrected for finite population size) - sampling method

Nevertheless, the different predictions differ not only in terms of precision. Sampling and predictive modeling actually result in substantially different predictions. Comparing the predicted *numbers* (rather than rates) of essential patents in different firms' portfolios, with predictions based on sampling (Figure 5) and predictive modeling (Figure 6), it is apparent that there are sizeable differences not only in the predicted numbers of essential firms in firms' portfolios, but even the ranking of the five largest portfolios of patents essential to 5G standards.

Figure 7 displays a scatter plot of the predicted shares of essential patents in different firms' portfolios, where the sampling rate is on the x-, and the rate predicted by the logistic regression model is on the y-axis. While there obviously is a positive correlation between the two different predictions, this correlation is surprisingly weak, and there are many observations for which the two methods yield inconsistent predictions. The scatter plot also reveals that predicted essentiality rates based on the model are far less distributed than the essentiality rates in the sample - either because the model predictions understate the true extent of variation between different firms, or extrapolations from very small samples yield overly extreme predictions of disparities between firms, or both.

In light of these disparities between different predictions, it is thus necessary to assess which method yields more accurate results. As we only have essentiality checks for the sample patents, we cannot directly assess the accuracy of the population predictions. We can however assess the out-of-performance of the two different methods, by using randomly drawn sub-samples of our sample.

5.2 Assessment of out-of-sample performance

In this section, we make predictions about essentiality shares in firm's portfolios in half of our sample, based on observations of the other half of our sample. This allows us to

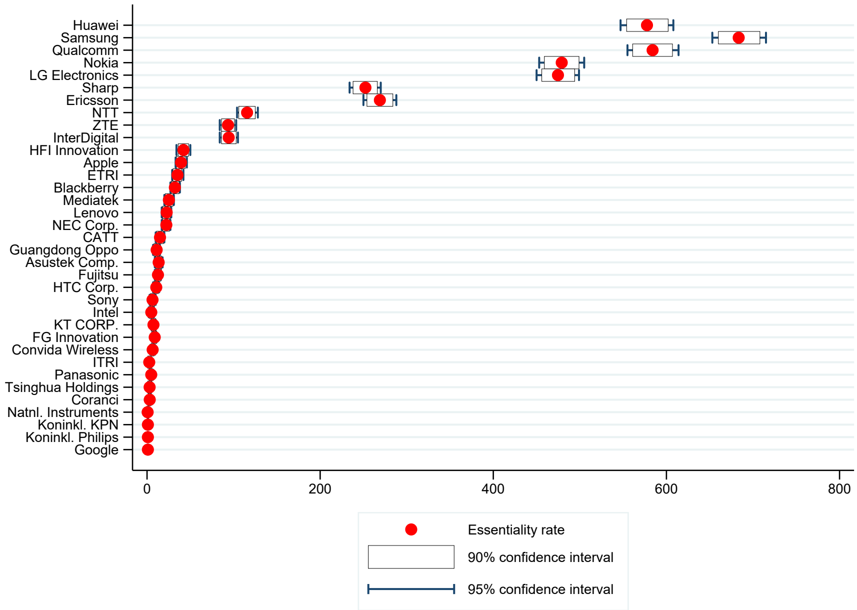

Figure 6: Number of actual SEPs and confidence intervals - logistic regression method

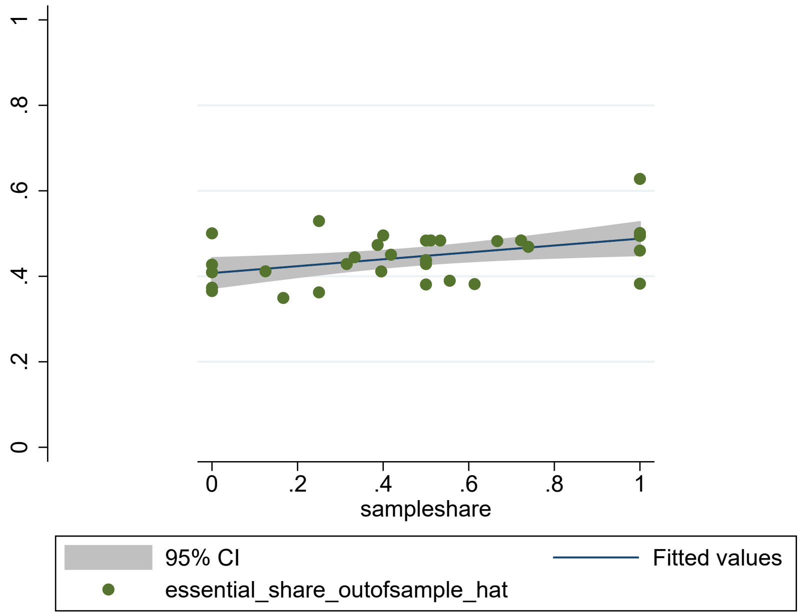

Figure 7: Correlation between predictions based on different methods

assess the capacity of different methods to correctly predict essentiality rates outside of the sample that was submitted to essentiality checks. Specifically, we make 200 random partitions of our sample, and aggregate the results from our assessments into different aggregate measures of the methods' accuracy. In particular, we calculate *average error*, *total bias*, and *firm level bias*.

We predict out-of-sample numbers of essential patents for each firm $i \in N$, for each sample partition $j \in \{1, 2, ..., 200\}$. We denote $\epsilon_{i,j} = \hat{x_{i,j}} - x_{i,j}$ the individual prediction error per firm and sample, i.e. the difference between the predicted number of essential patents of firm i in the excluded subsample j, and the "true" number of essential patents (as per our essentiality checks) of firm i in the excluded subsample j.

Average error is simply the average *absolute value* of all individual errors in the assessment of numbers of essential patents per firm, i.e. $\overline{|\epsilon_{i,j}|}$, and *total bias* is the arithmetic mean of all individual errors $\bar{\epsilon_{i,j}}$. If a method e.g. misses the "true" number of a firm's essential patents on average by one patent, but is just as like to over- as to under-estimate the true number, the average error is 1, and total bias is 0. Positive total bias indicates that the method generally over-, and negative bias indicates that it generally under-estimates the true number of essential patents in firms' portfolios (on average, across all firms). We define *firm level bias* as the the sample average of the absolute values of the arithmetic means of individual assessment errors per firm, or

$$\frac{\sum_{i=1}^{N} |\frac{\sum_{j=1}^{200} \epsilon_{i,j}}{200}|}{N}$$

If a method *systematically* over- or under-estimates the numbers of essential patents in a particular firm's portfolio, the average of assessment errors over sample partitions for that particular firm will differ from zero. The average amount of these firm level discrepancies from the mean zero assessment error baseline is a measure of the method's overall "fairness", i.e. its propensity to systematically favor or disadvantage different firms.

In Figure 8, we present *average errors* in the predictions of out-of-sample essentiality rates for the 14 largest portfolios in our sample. The figure represents the average error as a percentage of the true number of essential patents in the out-of-sample portion of the portfolio, for different methods.[33] We can see that average sampling error increases as portfolio size decreases - sampling is less reliable for smaller portfolios. The average error of regression approaches (either direct logit regression using patent-level characteristics, or two-stage predictive modeling aggregating patent-TS-level information) is relatively independent of the size of the portfolio,. Therefore, regression-based approaches are substantially more accurate than sampling for smaller portfolios.

Nevertheless, it is important to distinguish between random noise and systematic error (bias). In Figure 9, we present the *distribution* of assessment errors based on the sampling method, for the largest 14 portfolios. Once again, we can see that the accuracy

[33]In this preliminary version, the scale does not correspond to the true magnitude of average error. To be specific, Figure 8 depicts the average of the squares of assessement errors across sample partitions, divided by the true number of essential patents in the excluded portion of the portfolio.

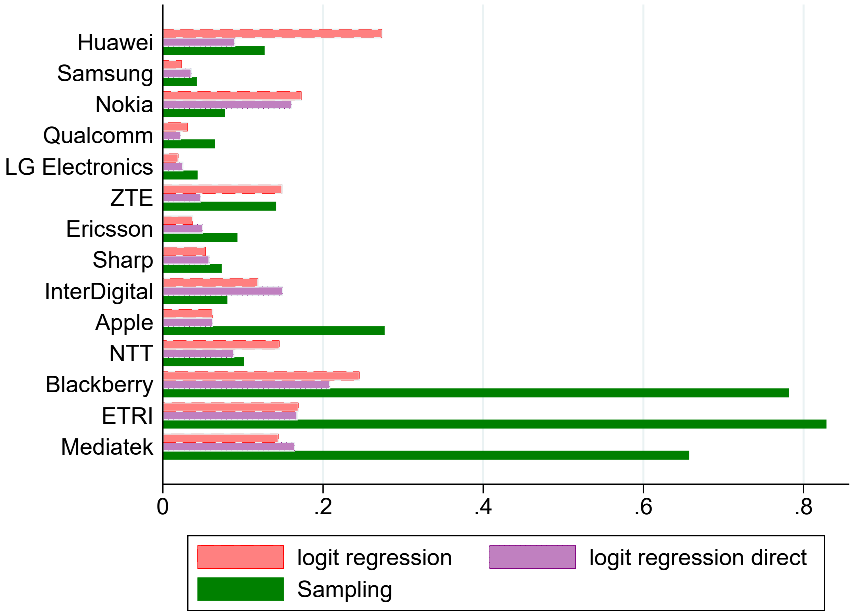

Figure 8: Average magnitude of prediction errors of different methods - 14 largest portfolios

of the method decreases when sample size decreases - the range of assessment errors becomes increasingly large. Nevertheless, the mean assessment error is always close to zero, confirming that sampling is unbiased - on average, sampling is just as likely to over- as it is to under-estimate the number of truly essential patents out-of-sample, for every individual firm.[34]

Replicating the same analysis with the predictions based on (direct) logistic regression in the sample yields a very different picture. While the range of assessment errors does not increase for smaller portfolios, and the absolute values of assessment errors for smaller portfolios are thus smaller using this method, the mean prediction error by firm may be very substantially (and statistically significantly) different from zero. This means that for some firms (Samsung, Sharp, Interdigital, NTT), our logistic regression approach systematically under-estimates the out-of-sample number of truly essential patents, whereas for other firms (in particular Nokia), it systematically over-estimates that number.

When comparing the performance of sampling and predictive modeling, there is thus an inherend trade-off between precision and bias - predicting the number of essential patents based on observable patent characteristics may improve upon sampling in terms of average accuracy (especially for smaller portfolios), but only sampling is inherently (axiomatically) unbiased. We will discuss the practical implications of this trade-off in the Discussion section below.

Empirically, we can attempt to resolve the trade-off by combining elements of sampling and modeling. One such combination consists in introducing firm-fixed effects in the logistic regression predicting the probability of an individual patent to be essential (Model 3). Similar to sampling, this approach uses observations of a firm's essentiality rate in

[34]This is of course axiomatically true for any random sampling.

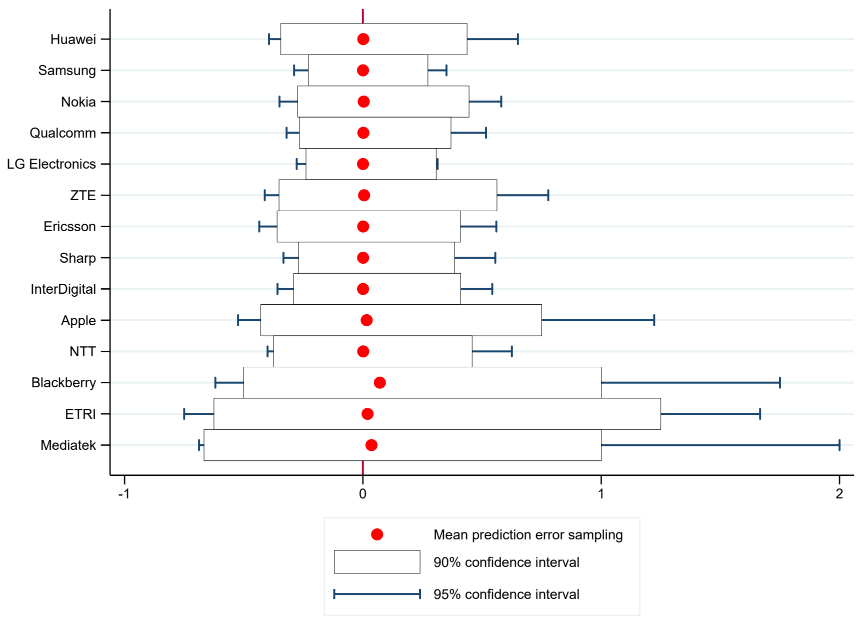

Figure 9: Prediction error and bias based on sampling method

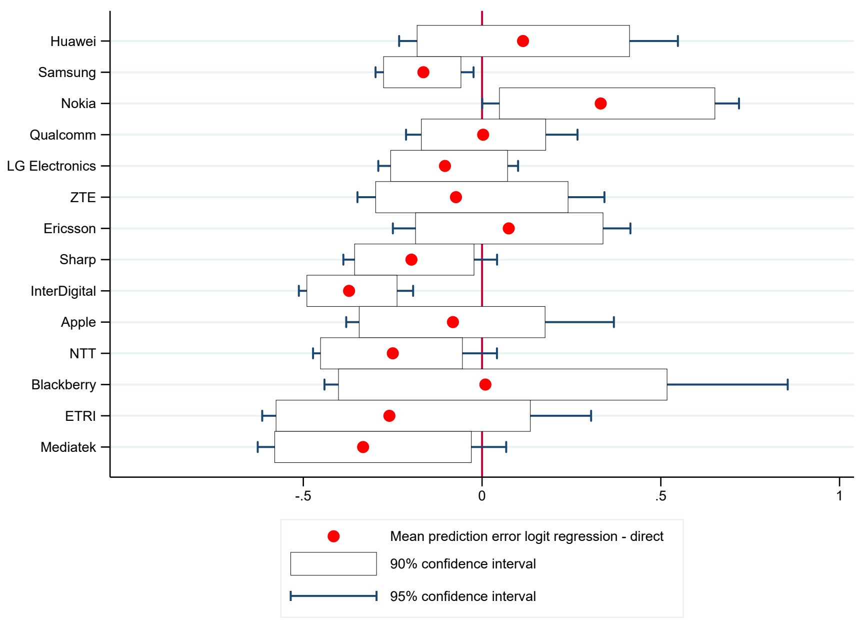

Figure 10: Prediction error and bias based on direct logistic regression method

the sample to predict the same firm's essentiality rate out-of-sample, but it combines this information with information about other covariates.

We present the overall out-of-sample performance measures of our different approaches in Table 11. The table confirms that logistic regression produces a smaller average error in the assessment of firms' out-of-sample essentiality rates than sampling. The advantage of modeling is relatively modest - while sampling produces an average error of 1.75 patents (or an average error of 20% of the true number of out-of-sample essential patents), the best logistic model produces an average error of 1.61 patents (or 17.8% of the true number of out-of-sample essential patents). Sampling is just as accurate as the best logistic regression model in predicting essentiality rates in the largest ten portfolios (both methods produce an average of approx. 3 patents on average, which - given the larger size of the portfolios - represents an average error of 10% of the true number), but predictive modeling more significantly outperforms sampling for the remaining portfolios (24% instead of 29% average error).

Nevertheless, Table 11 also confirms that all logistic regression appraoches are subject to bias. The amount of *total* bias seems relatively acceptable - on average, different logistic regression models over- or under-estimate the true number of essential patents in a firm's portfolio by 1 to 4% (i.e. some models are systematically overly optimistic, whereas others are systematically overly pessimistic, but this type of bias is not very pronounced). More worryingly, all models produce significant firm-level bias (i.e. systematically over-estimate the essentiality ratios of some firms, and systematically under-estimate the ratios of other firms). The different models produce a systematic error of 14 up to 18% of the true number of essential patents (whereas the risk of systematic error with sampling is negligible). The results imply that the largest part of the assessment error using logistic regression is systematic error, especially in larger portfolios.[35]

		average error			*total bias*			*firm-level bias*		
		All	top10	others	All	top10	others	All	top10	others
sampling	N	1.749	3.069	0.864	0.0280	0.0788	-0.0061	0.0839	0.146	0.0513
	pct	0.206	0.100	0.292	0.0017	0.0024	0.0012	0.0121	0.0044	0.0177
M1: logit (2 stage)	N	1.718	3.120	0.776	0.0817	0.416	-0.143	1.262	2.601	0.558
	pct	0.186	0.0998	0.256	0.0062	-0.0189	0.0265	0.155	0.0809	0.208
M2: logit (direct)	N	1.606	2.937	0.713	-0.472	-0.861	-0.211	1.092	2.232	0.493
	pct	0.178	0.100	0.240	-0.0254	-0.0528	-0.0033	0.136	0.0797	0.176
M3: logit (direct) FEs	N	1.859	3.376	0.841	-0.531	-0.867	-0.306	1.401	2.828	0.650
	pct	0.202	0.119	0.270	-0.0404	-0.0673	-0.0186	0.176	0.102	0.228

Table 11: Comparison of different methods - average precision and bias

[35] e.g. direct logistic regression produces an average assessment error of 10% in the largest portfolios, where 8% are attributable to systematic error (firm-level bias). Sampling also produces an average assessment error of 10% for these ten portfolios, but this error is entirely random - in our 200 random sample partitions, we find less than 1% firm-level bias with sampling (and axiomatically, we know that firm-level bias with sampling converges to zero).

6 Assessment uncertainty

So far, we have compared two different methods of extrapolating information about essentiality rates from a randomly drawn sample. As discussed, another alternative approach is to assess the entire population of declared SEPs. Given resource constraints on the assessment, that necessarily means that fewer resources are available to the assessment of each individual patent. By and large, it is plausible that more cursory assessments lead to a larger number of assessment errors. When comparing sample-based methods to assessments of the total population, it is thus necessary to analyze the trade-off between sampling and assessment errors.

So far, we have assessed the accuracy of different methods' predictions by assessing the method's ability of using a certain essentiality assessment in a (sub-)sample to predict the outcome of the same assessment out-of-sample. To assess to what extent these predicted essentiality rates correspond to the rates of patents that are "truly essential", one would need a sufficiently large sample of patents for which true essentiality is known with certainty. Nevertheless, no such sample exists. Even when individual SEPs are subjected to extensive expert assessments and counter-assessments, as for example in the context of complex litigation, the outcome of these determinations often continues to be open to challenges and debate. Empirically, there is no absolute measure of essentiality with which individual experts' assessments of essentiality could be compared.[36]

In order to analyze the consequences of uncertainty in the assessment of individual patents, we will use the tiered nature of our essentiality checks. So far, we have considered all patents that our experts classified as "fully mapped" or "partially mapped" (including "edge cases") to be "truly essential", resulting in an aggregate essentiality rate of 43.6% of the assessed patents. This assessment is certainly over-inclusive - in fact, our essentiality rate is situated at or above the upper end of what experts consider to be a plausible range of the share of truly essential patents among declared SEPs.[37]

On the other hand, considering that only "fully mapped" patents are essential would most likely be under-inclusive - by comparison to other existing studies, an essentiality ratio of only 12.4% seems excessively low. A plausible interpretation of the results of our essentiality checks is that "partially mapped" indicates a probabilistic statement about essentiality - the 12.4% "fully mapped" patents are very likely to be essential, the 56.4% "not mappable" patents are very unlikely to be essential, and there is significant residual uncertainty regarding the essentiality of the remaining 31.2% "partially mapped" patents. While there obviously also is uncertainty regarding the actual essentiality of "fully mapped" and "not mappable" patents, we are confident that the extent of uncertainty is greater among the "partially mapped" patents. We can thus use these tiers for a general analysis of the distribution of assessment uncertainty.

If different experts find vastly different shares of a population of declared SEPs to be essential, that necessarily means that their assessments of a large number of individual

[36]For a more detailed discussion of this point, see e.g. Mallinson (2021).

[37]Based on various studies of essentiality rates among declared SEPs, the contribution of the SEPs-Expert-Group (2021) e.g. notes that "an average essentiality ratio somewhere between 25% and 40% seems realistic, with substantial variation between standards and portfolios."

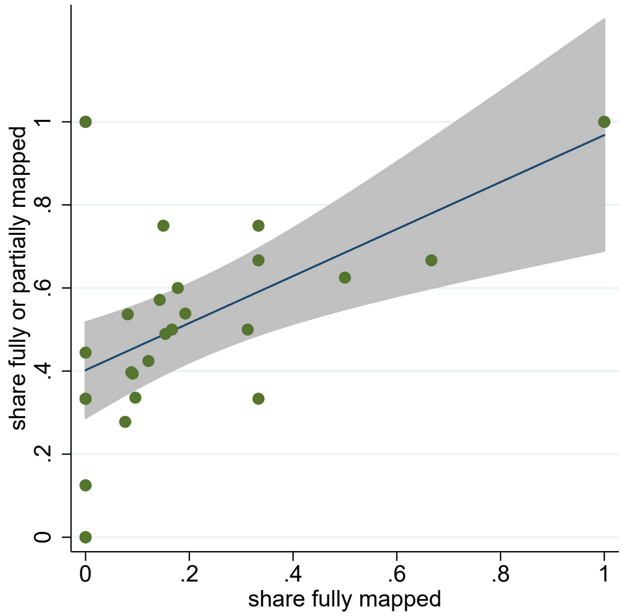

Figure 11: Scatter plot - essentiality shares of portfolios, optimistic vs pessimistic assessment

patents diverge. Nevertheless, the fact that experts disagree on the essentiality of individual patents must not necessarily mean that they disagree on the ranking of different portfolios. The question whether different experts' opinions on the essentiality of individual patents provides a reliable basis for the assessment of essentiality ratios in different portfolios hinges not as much on the number of times they get it right or wrong, as it depends on whether the errors they make are random noise or systematically benefit one firm over another. One (but not the only) potential source of systematic error arises if assessment uncertainty is distributed differently over different firms' portfolios.

For illustration, Figure 11 plots the shares of firms' patents that are "fully mapped" against the share of firms' patents that are either "fully" or "partially mapped". Following our general argument, a conservative (or "pessimistic") assessor would only consider the patents that are "fully mapped" to be essential; whereas an "optimistic" (or lenient) assessor would find all patents that are either partially or fully mapped to be essential. As can be seen from Figure 11, the judgments of pessimistic and optimistic assessors about essentiality shares in different firms' portfolios are clearly correlated - while the optimistic assessor finds much larger number of patents to be essential, portfolios that rank highest in an optimistic assessments also tend to rank highest in pessimistic assessments.

Nevertheless, the correlation is far from perfect. Some firms' portfolios include much larger shares of edge cases than others. Applying the sample shares of essential patents to the populations of declared SEPs, a conservative (pessimistic) assessor wiould find that Nokia holds the second, and LG Electronics the fourth largest number of actually essential 5G families; an optimistic (lenient) assessor however would find that LG Electronics ranks second, and Nokia only fourth. To compare how similar the predictions of the "optimistic" and the "pessimistic" assessor are, we compute Spearman rank correlations between the predicted numbers of essential patents in different firms' portfolios.

	fully mapped	fully or partially mapped	declared
fully mapped	1.0000		
fully or partially mapped	0.8825	1.0000	
declared	0.8153	0.9299	1.0000

Table 12: Spearman rank correlations - ranking of portfolios with pessimistic assessment, optimistic assessment, no assessment

The ranking of firm portfolios of the lenient examiner is actually more similar to a ranking based on simple counts of declared SEPs than to the conservative examiner's ranking. It is thus clear that assessment uncertainty regarding the essentiality of individual patents translates into significant uncertainty regarding essentiality ratios in different firms' portfolios - so much so that different experts' assessments may diverge more widely from each other than from a completely agnostic count of declared SEPs, without any essentiality assessment.

To understand the consequences of this assessment uncertainty, we simulate a "true" number of essential patents, which is in-between the views of the optimistic and the pessimistic assessor. We suppose (for the sake of exposition) that each partially mapped patent has a 50% likelihood of being essential; and simulated 200 realizations of the sample patents' actual essentiality.

For each of these hypothetical realizations, we simulated different types of assessment error. In a "light-touch" review, the examiner has no information about the essentiality of partially mapped patents. An optimistic assessor will classify each of these patents to be essential, a pessimistic assessor will classify none of these to be essential, and a neutral examiner will find each patent to be essential with a probability of 0.5.[38] Each of these assessors correctly classifies fully mapped and non-mappable patents, and - on average - for partially mapped patents, they all get it right half of the times, on average. In total, they correctly assess 84% of the patents in the sample.[39]

While the three examiners make exactly the same number of mistakes on individual patents, the errors that they make in the assessment of different firms' portfolios is different. Figure 12 depicts the average magnitude and direction of different examiners' errors in predicting the share of the total number of essential patents owned by any individual firm. While the optimistic expert obviously over-, and the pessimistic expert under-states the true number of essential patents in each firm's portfolio, they also make systematic firm-level errors - they systematically over-estimate the share of the total number of true essential patents that is held by some firms, at the expense of other firms. The neutral expert, on the other hand, is unbiased - not only with respect to the total share of essential patents, but also with respect to the shares of true essential patents owned by different firms.

This finding carries an important insight for the conduct of essentiality assessments. At first glance it seems to be good practice to only count those patents that are found to be

[38]which we simulate in 200 different realizations for each of the 200 hypothetical sample realizations; we thus average over 40,000 different iterations

[39]Our hypothetical "light-touch" reviewers perform thus about as well as the examiners in the pilot project by Bekkers et al. (2020).

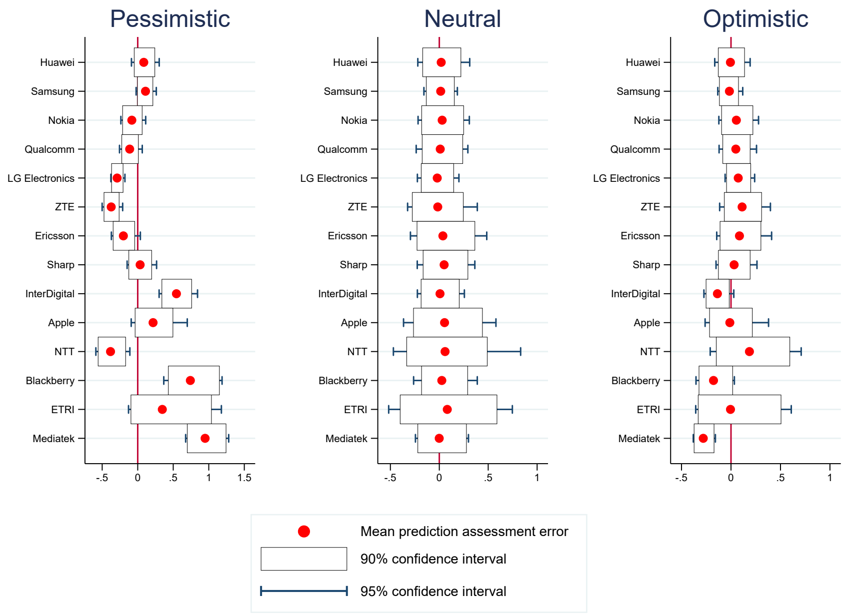

Figure 12: Average prediction error and bias - optimistic, neutral, and pessimistic assessors

essential with high degrees of confidence - or to only discard those declared SEPs that are quiet certainly not essential. Nevertheless, such assessment methods induce systematic bias between different firms. An unbiased review requires to adequately reflect the probabilistic nature of patents whose essentiality cannot be determined - if necessary with a coin flip, such as our hypothetical "neutral" assessor above.

Presumably, a more thorough review would produce more reliable information about patents' actual essentiality, including for difficult cases. We assume that after a thorough review, our examiner has a 75% chance of finding the true essentiality of a partially mapped patent.[40] In this more thorough assessment, all examiners (optimistic, pessimistic, or neutral) get it right 93% of the times, on average (for all patents). We can thus compare the outcome of a more thorough assessment of one (randomly drawn) half of the sample with a light-touch review of the entire sample.

		optimistic	pessimistic	neutral
thorough sample	average error	0.006	0.0088	0.0069
	bias	0.0014	0.0027	0.0006
light touch all patents	average error	0.0039	0.0084	0.0044
	bias	0.0022	0.0079	0.0005

Table 13: Comparison of thorough sample assessment and light touch review of all patents - error and bias

The results of that comparison are presented in Table 13. A "light-touch" assessment (accurate at 84%) of the entire sample generally produces a slightly lower average error in

[40]In a first step, the examiner has a 0.5 probability of discovering the essentiality of the patent. If the examiner does not discover the true essentiality, the pessimistic examiner finds the patent not to be essential, the optimistic examiner finds the patent to be essential, and a neutral examiner finds the patent to be essential with a random probability of 0.5 (each of these is correct half of the time).

the determination of firms' share in the total number of essential patents than a thorough review (accurate at 93%) in a randomly selected sub-sample, extrapolated to the entire sample. Nevertheless, when assessors are overly optimistic or pessimistic regarding the essentiality of declared SEPs, on average, a more thorough review of a smaller subsample is less prone to firm-level bias - while sampling produces slightly larger errors, on average, these errors are significantly less susceptible of systematically benefiting some firms at the expense of others.

Similar to our comparison between predictive modeling and sampling, we thus again encounter a trade-off between precision (the average absolute value of the error) and bias (a systematic error to the benefit of some firms and to the detriment of others). Depending on the composition of their portfolio, firms may know whether they are likely to benefit of a "pessimistic" or "optimistic" light touch review of all patents. This systematic error component is thus likely to lead to sustained controversies regarding the exact methodology of assessment. Sampling allows to increase the accuracy of any individual assessment, it thus trades systematic, predictable assessment error against random, unpredictable sampling error. While a firm may significantly benefit or lose from sampling error, *ex ante*, this error affects all firms the same way.

While there *can* thus be a trade-off between precision and bias in the choice between light-touch assessment of larger samples or populations and more thorough assessments of smaller subsamples, there are situations in which one method clearly dominates. For small average sample sizes (such as in our case, where we use 1,000 patents to determine essentiality ratios in 35 different portfolios), sampling error is large; a lighter-touch assessment of all patents with zero-centered assessment error may thus be the best method for this assessment. When sample sizes get larger, random sample error diminishes, whereas the systematic component (bias) of assessment errors remains - for larger portfolios (or larger sample sizes), thorough assessments within a sample are both more precise *and* less susceptible to bias than less thorough assessments of larger numbers of patents.

7 Conclusion

This article compares different methods of determining the share of different firms' declared SEPs that are actually essential. In particular, we have compared three viable approaches: sampling, predictive modeling, and light-touch reviews of all patents. We have derived two basic insights from our comparisons:

First, different methods are best suited to different contexts. While sampling is clearly the most suitable method to determine essentiality ratios in large patent portfolios, sampling-based estimates may become highly imprecise when sample size decreases. For an assessment of smaller populations and portfolios, a lighter assessment of every single patent can produce more accurate predictions than extrapolating the outcomes of more rigorous assessments in a random sample. For the assessments of essentiality ratios in larger numbers of smaller and medium size portfolios, predictive modeling can be an attractive solution. As predictive modeling can use the entire sample for making predictions about the out-of-sample essentiality ratios of each individual portfolio, it can make more accurate predictions than sampling alone, especially for the smaller portfolios in the population.

Second, the choice of a method involves a trade-off between bias and precision. While light-touch assessments of every patent and predictive modeling *can* - in some situations - produce smaller errors on average than sampling, they are prone to producing *systematic* error, i.e. bias that systematically favors some firms to the detriment of others. While sampling error may lead analysts to over- or under-represent the true share of essential patents in a particular firm's portfolio, each firm is a priori similarly likely to benefit or to lose. A verifiably random sample selection is thus inherently fair, and immune to manipulation.

The errors produced by predictive models and short-cuts in the examination of individual patents, on the other hand, are largely systematic errors. It must thus be assured that the process is carried out by parties that have no interest in biasing the outcome of the assessment. Furthermore, firms are likely to know (or to learn) which particular variant of each method is susceptible of favoring them. Similar to existing controversies over other aspects of FRAND determinations, it is thus plausible that different stakeholders will have entrenched preferences for specific SEP determination methods, with little hope for consensus. In that light, the (relatively modest) potential disadvantage of random sampling as compared to other methods in terms of precision may be a price worth paying in exchange of a prediction that is axiomatically unbiased and produces axiomatically accurate confidence intervals.

In addition to these fundamental insights, we have gained some insights into the best application of different methods. For predictive modeling, we have compared three different models. While our initially preferred patent-TS-specific model makes use of the greatest amount of information, and a model with firm fixed effects (effectively a hybrid between predictive modeling based on observable patent characteristics and pure sampling) has the best *in-sample* performance, a simpler direct logistic regression model (in line with the model proposed by Brachtendorf et al. (2020)) made the most accurate out-of-sample predictions. This underlines the importance of assessing out-of-sample performance of predictive models, and in particular their ability to predict the specific empirical magnitude of interest, rather than focusing on in-sample predictive power.

As for the uncertainty underlying the assessment of each individual patent, our results suggest that the reliability of these assessments does not primarily hinge on the rate at which different examiners agree on the essentiality of individual patents, but on the distribution of assessment errors over firm portfolios. In this regard, a patent count based on a probabilistic assessment of edge case patents is preferable to a "conservative" approach that only counts confirmed SEPs (or only discards those declared SEPs that are clearly not essential).

Overall, our results confirm that for the determination of essentiality shares in larger portfolios (and a fortiori the entire population of declared SEPs), thorough assessments of (possibly relatively small) randomly selected samples are the most accurate and unbiased method. This suggests that policy proposals that aim at evaluating the essentiality of every declared SEP would likely lead to an inefficient use of scarce examination resources.

We do not, however, take any position here on whether counts of declared SEPs, adjusted for "essentiality ratios", *should* be used at all for the determination of FRAND licensing terms. To be viable, such determinations would also need to account for the existing significant heterogeneity in the value of different SEPs. In that light, uncertainty regarding the *number* of truly essential patents in different firms' portfolios - sometimes portrayed to be among the most important sources of disagreements in SEP licensing - is perhaps one of the more easily solvable challenges in the valuation of SEP licenses.

References

J. Baron and H. Delcamp. The private and social value of patents in discrete and cumulative innovation. *Scientometrics*, 90(2):581–606, 2012.

J. Baron and T. Pohlmann. Mapping standards to patents using declarations of standard-essential patents. *Journal of Economics and Management Strategy*, 27(3):504–534, 2018.

R. Bekkers, C. Catalini, A. Martinelli, C. Righi, and T. Simcoe. Disclosure rules and declared essential patents. Technical report, National Bureau of Economic Research, 2017.

R. Bekkers, J. Henkel, E. M. Tur, T. van der Vorst, M. Driesse, B. Kang, A. Martinelli, W. Maas, B. Nijhof, E. Raiteri, et al. Pilot study for essentiality assessment of standard essential patents. *Launch event for Pilot Study for Essentiality Assessment of Standard Essential Patents*, 2020.

L. Brachtendorf, F. Gaessler, and D. Harhoff. Truly standard-essential patents? a semantics-based analysis. 2020.

R. Clancy, T. Eskildsen, N. Ruest, and J. Lin. Solr integration in the anserini information retrieval toolkit. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1285–1288, 2019.

D. E. Cooper. Evaluating standards essential patents in mobile cellular. *les Nouvelles-Journal of the Licensing Executives Society*, 54(4), 2019.

D. J. Goodman and R. A. Myers. 3g cellular standards and patents. In *2005 International Conference on Wireless Networks, Communications and Mobile Computing*, volume 1, pages 415–420. IEEE, 2005.

B. H. Hall, A. Jaffe, and M. Trajtenberg. Market value and patent citations. *RAND Journal of economics*, pages 16–38, 2005.

B. Kang and R. Bekkers. Just-in-time patents and the development of standards. *Research Policy*, 44(10):1948–1961, 2015.

B. Kang and K. Motohashi. Essential intellectual property rights and inventors' involvement in standardization. *Research Policy*, 44(2):483–492, March 2015.

J. O. Lanjouw and M. Schankerman. Patent quality and research productivity: Measuring innovation with multiple indicators. *The Economic Journal*, 114(495):441–465, 2004.

M. A. Lemley and T. Simcoe. How essential are standard-essential patents. *Cornell L. Rev.*, 104:607, 2018.

K. Mallinson. Essentiality rate inflation and random variability in sep counts with sampling and essentiality checking for top-down frand royalty rate setting. Technical report, WiseHarbor, 2021. available at https://www.wiseharbor.com/wp-content/uploads/2021/09/Perils-of-sampling-SEPs-Mallinson-30-Sept-2021.pdf.

A. C. Marco, J. D. Sarnoff, and A. Charles. Patent claims and patent scope. *Research Policy*, 48(9):103790, 2019.

V. K. Rangan and J. Yonamine. A hybrid machine learning with human assistance approach to analyzing standard essential patents. Technical report, google, 2021. unpublished working paper.

P. Regibeau, R. De Coninck, and H. Zenger. Transparency, predictability, and efficiency of sso-based standardization and sep licensing. Technical report, Report to the European Commission, 2016. https://ec.europa.eu/growth/content/study-transparency-predictability-and-efficiency-sso-based-standardization-and-sep-0_en.

M. Rysman and T. Simcoe. Patents and the performance of voluntary standard-setting organizations. *Management Science*, 54(11):1920–1934, November 2008.

SEPs-Expert-Group. Contribution to the debate on seps. Technical report, The European Commission's Group of Experts on Licensing and Valuation of Standard Essential Patents, 'SEP Expert Group', 2021. available at https://ec.europa.eu/docsroom/documents/45217.

R. Stitzing, P. Sääskilahti, J. Royer, and M. V. Audenrode. Over-declaration of standard essential patents and determinants of essentiality. *Available at SSRN 2951617*, 2017.

M. Trajtenberg. A Penny for Your Quotes: Patent Citations and the Value of Innovations. *RAND Journal of Economics*, 21(1):172–187, Spring 1990.

A Appendices

A.1 Appendix 1: Descriptive Statistics

assignee_norm	number_patents	number_families	number_families_sample
Samsung	3546	1985	139
Huawei	3210	1862	142
Qualcomm	2746	1535	98
Nokia	1992	1350	124
LG Electronics	2437	1238	118
Ericsson	1465	687	62
Sharp	1035	662	45
NTT	391	250	20
ZTE	301	227	64
InterDigital	647	207	23
HFI Innovation	265	112	9
ETRI	206	103	12
Apple	144	102	26
Mediatek	129	85	8
Blackberry	213	72	14
NEC Corp.	160	58	5
Guangdong Oppo	68	53	8
CATT	59	52	8
Lenovo	81	49	2
Asustek Comp.	43	39	3
Panasonic	65	30	3
HTC Corp.	35	26	3
Fujitsu	107	25	3
Sony	70	21	3
KT CORP.	26	20	3
Intel	21	17	3
FG Innovation	13	13	2
Tsinghua Holdings	12	11	1
Convida Wireless	12	10	3
ITRI	10	9	3
Coranci	6	5	2
Koninkl. Philips	4	4	1
Natnl. Instruments	7	3	1
Koninkl. KPN	5	2	1
Google	1	1	1
Total	19488	10860	960

Table 14: Descriptive statistics - number of patents and INPADOC patent families by assignee

assignee_norm	fullymapped	partially	notmappable	declared_ts_per_patent	ts_decl_and_mapped_per_pat
Samsung	0.157	0.321	0.521	2.264	0.536
Huawei	0.0699	0.203	0.727	6.979	0.510
Qualcomm	0.0918	0.296	0.612	3.194	0.520
Nokia	0.0960	0.240	0.664	1.480	0.336
LG Electronics	0.0847	0.458	0.458	2.517	0.619
Ericsson	0.109	0.297	0.594	1.406	0.438
Sharp	0.178	0.422	0.400	4.600	0.844
NTT	0.150	0.600	0.250	1.400	0.850
ZTE	0.0769	0.323	0.600	5	0.631
InterDigital	0.375	0.375	0.250	2.250	1.417
HFI Innovation	0	0.444	0.556	2	0.556
ETRI	0.167	0.333	0.500	5.667	0.917
Apple	0.192	0.346	0.462	3.192	0.769
Mediatek	0.444	0.111	0.444	1.333	0.444
Blackberry	0.313	0.250	0.438	3.313	1.188
NEC Corp.	0.200	0.400	0.400	2.800	0.600
Guangdong Oppo	0	0.125	0.875	1.500	0.375
CATT	0	0.125	0.875	4.250	0.125
Lenovo	1	0	0	1	0.500
Asustek Comp.	0	0.333	0.667	1.333	0.667
Panasonic	0.333	0.333	0.333	1	0.667
HTC Corp.	0.333	0	0.667	1.333	0.333
Fujitsu	0	0.333	0.667	3.667	1
Sony	0	0.333	0.667	1.667	0.667
KT CORP.	0.667	0	0.333	2.333	1
Intel	0	0	1	3.333	0
FG Innovation	1	0	0	2.500	1.500
Tsinghua Holdings	0	0	1	1	0
Convida Wireless	0	1	0	2.333	1.667
ITRI	0	0	1	2	0
Coranci	0	1	0	1	1
Koninkl. Philips	1	0	0	1	1
Natnl. Instruments	0	0	1	1	0
Koninkl. KPN	0	1	0	2	2
Google	0	1	0	4	0
Total	0.124	0.312	0.564	3.284	0.582

Table 15: Descriptive statistics - Mapping status of sample patents by assignee

A.2 Appendix 2: Semantic scoring methodology

In order to semantically compare declared patents to declared TS (technical specifications) a parsing algorithm indexes and separates all independent patent claims as well as all independent sections of each TS. The indexing is based on the Lucene based Solr index (Clancy et al., 2019). If the declaration does not indicate a version number of the TS to which the patent is declared to be potentially essential, the latest TS version (as of October 2020) is identified and considered.

Each distinct patent claim and each distinct TS sections are semantically compared. TS documents have 80-350 sections, patents have up to 5 independent claims. The word count of sections is on average 2.5 times higher compared to the word count in claims. For each declared patent TS combination, we identify between 400-1,750 patent claim-TS section combinations. Each combination is semantically compared, and given a semantic similarity score. We identify the highest-scoring patent TS combinations out of all possible patent claim TS section combinations, identifying the claim number, the section number, and the semantic score. The model used for calculating the score is the Latent Semantic Indexing (LSI) model, which follows a 5-step approach:

1. We use the different sections of each standard to be compared to worldwide independent patent claims for the textual input of the similarity analysis.
2. We create a word vector matrix (term-document matrix) where each row corresponds to a term (of the documents of interest), and each column corresponds to a document. Each element (m,n) in the matrix corresponds to the frequency that the term m occurs in document n. We apply Log Entropy as local and global term weighting.
3. Singular value composition (SVD) is used to reduce this matrix to a product of three matrices, one of which has non-zero values (the singular values) only on the diagonal.
4. Dimensionality is reduced by deleting all but the k largest values on this diagonal, together with the corresponding columns in the other two matrices. This truncation process is used to generate a k-dimensional vector space. Both terms and documents are represented by k-dimensional vectors in this vector space.
5. The relatedness of any two objects represented in the space is reflected by the proximity of their representation vectors, in our case: cosine measure.

Latent semantic indexing, is a class of techniques where documents are represented as vectors in term space. The semantic analysis of a corpus is the task of building structures that approximate concepts from a large set of documents. The LSI model produces semantic similarity scores for each patent claim and technical specification section combination. Scores are presented in percentages of similarities.