Measuring Patent Assessment Quality

-

Analyzing the Degree and Kind of (In)Consistency in Patent Offices' Decision Making

This version: June 2007

Paul F. Burke[1,4]
Markus Reitzig [2,3,4]

Abstract:

We argue that consistent decision making in judging a patent's validity and basing this on its underlying technological quality are important elements of patent office service ("assessment") quality. To understand which level of assessment quality patent offices (can) provide, particularly in new technological areas, we study the concordance of the European Patent Office's (EPO) granting and opposition decisions for individual patents. Using the biotechnology industry in the 1980s (an emerging patenting area then) as an example, we find no empirical evidence that the EPO provided maximal or optimal assessment quality as far as can be told from bibliographic indicators. We discuss research limitations and consequences of this first empirical analysis, and suggest ideas for refinements in future work.

Keywords: Patents, quality, novelty, inventive step, discrete choice, variance decomposition

JEL-Classifications: C25, C51, K41, L00, L20

[1] The University of Technology, Sydney, PO Box 123, Broadway NSW 2007/Australia. paul.burke@uts.edu.au.

[2] London Business School, Department of Strategic and International Management, Sussex Place, Regent's Park, London NW1 4SA. Mreitzig@london.edu. Phone: +44 (0)20 7000 8714 (Miss Claire Lymer). Fax: +44 (0)20 7000 8701.

[3] Corresponding author

[4] *RIPE* Network for the Research in Intellectual Property Economics

1 INTRODUCTION

This paper is motivated by the ongoing discussion in the popular media, among policy makers and in academic circles (e.g., Samuelson, 2004) about "decreasing patent quality" as a consequence of decreasing patent office "service quality". "Patent quality" can be defined along two major dimensions; namely (a) the techno-(economic) quality created by a patent's underlying invention; and (b) the legal quality created by a patent's reliability as an enforceable property right (Thomas, 2002). This second (legal) dimension of quality pertains to a patent's sustainability when challenged. In this regard, a patent office's role in providing consistent service is twofold. First, it must consistently assess patent applications and only grant patents with sufficient technological quality (novelty/inventive step) (Reitzig, 2005). Second, it must assess those patents that have been challenged using criteria that are consistent with initial assessments. The quality of patent office assessments, however, has been heavily debated. Critics claim that assessment quality is decreasing. Many of these claims are based on the observation that patent offices are (allegedly) *incapable* of examining patent quality in areas where no prior art patents exist. This concern is most applicable in emerging technological areas, such as software or nanotechnology (Merges, 1999). However, related empirical evidence is mixed (Allison & Tiller, 2003; Merges, 1999; Quillen & Webster, 2001). The impact that patent validity assessments has on national innovative activity and the heated debate at present suggests that there exists the need for robust empirical evidence shedding light on the question of how good patent granting procedures are. At present, it is unclear how reliable it is that a patent, once granted, will survive challenges (i.e., validity suits), and if not, the reasons why such inconsistency occurs.

Our research focuses on one core component of "assessment quality", validity examination quality, and whether patent offices (can) provide it in novel technological areas. We empirically assess whether the European Patent Office (EPO) consistently based its judgments on technological quality when repeatedly deciding on the validity of identical individual patents during granting and challenge phases. In addition, we critically discuss how such consistency can be *undesirable* in ensuring assessment quality. We focus our empirical analysis to study the existence of systematic

(in)consistencies between these granting decisions and challenge decisions relating to individual patents.

We focus on the EPO for various reasons. First, the EPO is relevant in being one of the biggest patent offices worldwide, administering the protection of inventions for the largest internal global market. Second – and an important reason from our research perspective – EPO patent application data allows one to observe granting, "challenge" (i.e. opposition), and opposition outcomes. For the purpose of this paper, we assume European "oppositions" are similar to U.S. "central validity suits" (see Harhoff & Reitzig, 2004 for a detailed comparison), such that any EPO ruling on an opposition will affect all designated national member states where applicants sought patent protection. To test our model, we require observations of opposition rulings made by patent offices; subsequently, we focus on biotechnology patents initially filed between 1978 and 1987. At this time, biotechnology was an emerging technology from the perspective of the EPO (Orsigeno, 1989). In this regard, decisions by patent offices about biotech patents are comparable to decisions currently relating to software and nanotechnology, but for which challenge decision data are not yet available (Merges, 1999).

Our econometric analysis and theoretical approach to questions of assessment consistency provides some notable departures. We model information on the technological quality of patents available to the EPO by using information in the form of standard bibliographic indicators (e.g., backward references; family size; forward cites). We estimate a series of discrete choice models using these bibliographic indicators to explain patent grant and opposition outcomes. The parameter estimates provide insight (termed "paramorphic representation" by Hoffman, 1960) into how patent offices make assessments; namely the estimates quantify those dimensions of quality that are given more emphasis in patent office decisions. If patent offices act consistently, we expect *similar* parameter estimates for granting and opposition decisions. A major econometric and theoretical obstacle, however, is that these two sets of estimates are confounded with decision variability, as well as other sources of random error. Our econometric approach is based upon and extends a set of discrete choice models already used in other fields, such as marketing and transportation (see Louviere, Hensher and Swait, 2000; Louviere, 2001). We apply these models here because they can

account for issues of variability – including its confounding effects on parameter estimates – ensuring valid comparisons and conclusions.

We summarize our major empirical findings as follows: within the general limitations of our indicator-based research design, particularly our framework of assumptions and for our chosen data, we cannot confirm that the EPO assessed the technological quality of biotech patents consistently in the early 1980s. Specifically, we do not find convincing evidence of congruent systematic effects driving the assessment of a patent's technological quality during granting or opposition stages. We consider that examiners (who decide initially whether to grant a patent) and members of opposition divisions (who decide on cases when a granted patent is opposed) may have different information. We use preliminary bibliographic indicators to capture informational change between grant and opposition stages. After including these indicators in our models, we find that examiners and opposition divisions do not primarily disagree because one holds more "costly-to-generate" information than the other. Rather, it seems that patent offices assess patent-quality related information differently at these two stages; hence, it less likely that inconsistency is driven by an effort of patent offices to allocate resources *optimally* between grant and opposition. Time effects, which do not relate to a patents' technological quality, seem to explain most of the decision-making of patent offices across stages. These findings indicate that changing environmental conditions in which examiners operate, as well as their own learning over time, account for most of the decision making pattern.

In Section 2 we provide clearer definitions of the term "patent quality", conceptually link it to the central notion of "assessment quality", and discuss prior research. In Section 3 we develop our testable hypotheses and introduce our methodology. In Section 4 we describe the data and present estimation results. In Section 5 we discuss these and highlight several research limitations. Section 6 concludes and presents new questions for research into patent quality.

2 BACKGROUND – THEORY AND RELATED PRIOR EMPIRICS

2.1 Theory

2.1.1 Dimensions of "patent quality" – techno-economic and legal considerations

The economic arguments for patent protection relate to stimulating research and development (R&D); encouraging disclosure of technological knowledge; and, facilitating technology transfer (see Gallini, 2001). Underlying these arguments are two fundamental assumptions: (1) technology contributes to social welfare; and (2) the economic value of a patent is in its underlying technological sophistication (for critiques on these assumptions see Merges, 1988, and Reitzig, 2005). Prior literature suggests that patent offices should adjust their minimal conditions for patentability requirements to guarantee that inventions have a *sufficient* level of technological quality (Nordhaus, 1967; Scotchmer & Green, 1990; Green & Scotchmer, 1995; Barton, 2001). This is because many economists traditionally assume that technological quality correlates highly with a patent's economic value; therefore it is also termed *technological 'merit'* (see Merges, 1988). This is also why we use the terms "techno-economic" and "technological quality" interchangeably in the following. Economists believe that a patent must exceed an *absolute threshold of technological quality* for it to be granted.

Lawyers, on the other hand, traditionally concentrate on another dimension of a patent's quality: namely, its legal sustainability. According to Thomas (2002)

> *"'quality patents' are [...] valid patents [which may] be reliably enforced in court, consistently expected to surmount validity challenges, and dependably employed as a technology transfer tool"* (Thomas, 2002: 730)

This definition focuses exclusively on legal *certainty* or *consistency*. For Thomas, as for other lawyers, the optimal absolute adjustment of patentability parameters (disclosure, novelty, and inventive step), and identifying a technological threshold appear second-order compared with ensuring requirements of legal certainty; hence, lawyers focus on the *relative comparability* of patent assessments from one case to another. In general, Merges (1999) appears to share Thomas' (2002) standpoint, however, he admits that resource constraints during the granting procedure prohibit calling every patent a bad patent that does not surmount a validity challenge after being granted:

"A 'bad patent' is a patent that should have been weeded out after a reasonable investment of effort, but was not" (Merges, 1999: 581)

2.1.2 The link between patent quality and patent assessment quality

Arguably, assessing patent validity at various stages of a patent (application)'s 'life' is the most important *service* that patent offices provide. Incorporating work by Parasuraman et al. (1991), two major constructs constitute the quality of a patent office's services; namely the ability of patent offices to act in a *consumer orientated* fashion and to do so in a *reliable* manner. In the context of patent offices, there are various customers; namely, these are the applicants, as well as society as a whole. These stakeholders expect patent offices to judge patentability requirements *correctly* against a given yardstick (i.e., economic dimension of quality), and *reliably* (or consistently) in the sense that the service can be trusted (i.e., legal dimension of quality). For the purpose of this paper, and, in keeping with prior literature, we therefore define patent assessment quality as:

"a patent office's consistent categorization of patents along a dimension of technological quality leading to sustainable property rights".

2.1.3 Maximal and optimal levels of patent assessment quality

In an *ideal* world where patent offices can objectively assess patent quality and maximize it, we would expect to observe no inconsistencies in patent validity assessments. That is, once a patent is granted, we would *not* expect it to be amended or revoked/annulled. Service quality, however, comes at a cost, and as such is not a self-purpose but will have a theoretical optimum. In particular, patent assessments are the product of a (subjective) human activity, prone to errors arising from our cognitive and emotional imperfections. Completely eliminating these errors comes at a prohibitive cost. Thus, various errors occur that force us to rethink the desirability of "consistent" assessments[5]. We now discuss three errors in respect to the initial granting procedure (see Box 1).

Insert Box 1 about here

[5] The source of such imperfections (e.g., demand to meet quotas) or how they can be minimised in general (e.g., training) is not the purpose of the paper and we do not discuss these in any detail. An assumption is made that patents are assessed in an equitable fashion, given the environment encountered by assessors.

First, patent offices may make minor errors in their initial subjective judgment of the objective technological quality of patents (Stage 1), although this may be inconsequential to decision outcomes (we label these error "type α"). That is, even allowing for some human error between objective and subjective assessments of quality, examiners may *still* correctly classify a good patent as good and a bad patent as bad. Second, patent offices may incorrectly reject patents fulfilling patentability requirements (we label these error "type β"). Third, patent offices may incorrectly grant a patent, even though it has an objectively poor level of quality (we label this error "type γ").[6] The latter two errors are most consequential for patent applicants.

Patent offices can rectify initial patent granting errors, however, provided that they have mechanisms in place for correcting these. At the EPO, the "appeal to grant" procedure corrects for "type β" errors, and the "opposition procedure" corrects for "type γ" errors. Patent offices must optimally balance the various errors that occur. This is difficult if one considers that some errors in patent assessment may be desirable if offices are willing to rely on re-assessment procedures to rectify their mistakes in their quest to maximize *ex-post* efficiency.

For example, being more stringent in Stage 1 (granting) assessments (indicated by 'Threshold A' in Box 1) and granting *fewer* patents initially, places greater demands on patent office resources in managing the procedures of "appeal to grant". Demands may also fall on inventors themselves, as it delays patentable inventions and requires allocating resources to rectify mistakes through the appeals to grant procedures. This could increase uncertainty and delay investment in emerging industries. Indeed, some patents may never see the light of day as a result of stringent granting policies. On the other hand, the use of *less* stringent thresholds in Stage 1 assessments (indicated by 'Threshold B' in Box 1) implies a greater level of γ errors. In practice, a greater number of poor patents may saturate the market and stakeholders may lose faith in the granting procedures of patent offices.

Similar types of errors exist at Stage 3, where previously granted patents are reassessed upon challenge. The major difference, however, is that there may exist opportunities to minimize such errors, given the privilege of hindsight and that reviewers have extra search resources for prior art.

[6] We do not enter the philosophical discussion whether "objective" technological quality exists, but

That is, in Stage 3, there will still be some level of misclassification of patents in terms of patent quality, but we anticipate that the level of such errors will be lower relative to Stage 1; hence, assessments in Stage 3 should be closer to the "true" objective assessment of technological quality relative to initial assessments in Stage 1. Indeed, this paper is motivated by accounting for these anticipated error differences across stages of assessment. Once accounted for, as is achieved in our subsequent models presented, questions of systematic consistency can then be properly examined.

These "desirable errors" should largely be exceptions to the (empirical) rule; otherwise, there is a danger these errors could become standard, resulting in a completely arbitrary granting process for patents. This can hardly be optimal. On average, therefore, validity decisions should be systematically consistent. That is, while we make little comment on *how* patent offices can reduce errors, we conjecture that it is desirable to improve the human imperfections through various strategies such as additional examiner training or an appropriate number of examiners to meet demand.

2.2 Prior empirical research into patent assessment quality

To the best of our knowledge, few large-scale empirical studies shed light on the question of what level of assessment quality patent offices (can) provide, particularly under the aggravating circumstances prevalent in novel technological areas. Patent offices face a series of reproaches suggesting that low 'patent quality' is becoming a problem. We classify these into two major categories. Namely, these categories capture (a) commentaries or case-based annotations, stating that inventive step (US: non-obviousness) requirements fell systematically over time, and granted patents are falling below an acceptable threshold (Samuelson, 2004); as well as (b) comments and anecdotal observations about the (increasing) assessment inconsistency by patent offices with respect to the technological quality of comparably sophisticated inventions submitted at the same time (Merges, 1999: p. 589). We summarise the few large-scale empirical findings speaking to these allegations in the following sub section.

2.2.1 Decreasing technological quality requirements over time

Sampat et al. (2003) track changes in patents' technological quality using forward citation measures. For their chosen sample of university patents, the authors observe a significant decline in

rather, in line with the whole idea of a patent system, take this assumption for granted.

forward citations since 1980. It is unclear, however, whether one can generalise these results to non-university patents. Moreover, it is unclear whether this decline in technological quality was because patentability standards for non-obviousness steps were actually falling.

Sanyal & Jaffe (2005) examine several effects that may account for observable increases in patenting rates. In particular, they attempt to disentangle the effects relating to potential decreasing patentability standards. Within their framework of assumptions, they attribute increases in filing rates to increases in overall inventiveness, whereas evidence for decreases in patenting standards is mixed.

2.2.2 Inconsistent assessments of technological quality for similar patents

Quillen & Webster (2001) examine the USPTO (United States Patent Office) granting rates. The authors indicate that split applications (continuations) have a higher chance of being granted than original applications – all else being equal – because of internal procedures. Indirectly, their results support the hypothesis that patent quality was judged along different dimensions in similar cases. For a selection of patents with identical priorities (i.e., a patent stems from the same original invention), or 'twins', Graham et al. (2002) track the fate of European oppositions and compare them to US re-examinations. Their findings indicate that the European and US offices rule distinctly differently in similar cases. Although we recognize the attractiveness of their approach, their focus on inter-institutional consistency contrasts with our focus on intra-institutional patent assessment consistency.

Allison & Tiller (2003) compare U.S. business method patents with those from 'established' patenting areas along a series of bibliographic indicators. Within their framework of assumptions, and, depending on each indicator's ability to capture effects relating to patent quality, the authors find no significant differences between business method patents and other patents. Despite the study's contribution and although it uses a rich underlying data base, the rather descriptive character of its results may render it difficult to draw final conclusions. In essence, the authors compare patents using several bibliographic indicators without forcing these indicators to capture quality related phenomena.[7]

3 EMPIRICAL RESEARCH DESIGN

3.1 A Research Gap of Public Sector Relevance

As the previous section highlighted, large-scale empirical data on patent offices' *consistency of discriminating between patentable and non-patentable technology along a yardstick of technological quality* is missing. From scientific and applied perspectives, this research gap calls for an answer. The public sector relevance of this research gap appears tremendous. Reactions of patent offices to allegations that their service quality is sub-standard reflect this.[8]

Although a fully-fledged analysis of EPO service quality would examine α, β, and γ errors, we focus on analyzing γ (and somewhat less on α) errors; that is, we seek to understand patent office "service quality" by analyzing if and why patents were "falsely" granted.[9] We do this in two ways. First, we require an understanding of the *degree* of systematic (in)consistency for the reasons mentioned before (see 2.1.3). We wish to determine if systematic inconsistency are exceptions, since neither a maximal nor optimal degree of assessment quality can exist if judgment inconsistencies are the norm. Second, we focus on understanding the residual inconsistency: we consider reasons *why* patent assessments are (in)consistent and what this reveals about patent offices' resource allocations to patent assessment

To examine these two questions, and assessing the EPO's assessment quality when it is exposed to the challenge of judging new technologies (Merges, 1999), we first map the technological quality of a patent econometrically. We then compare these assessments during the granting phase and the subsequent challenge phase for a given set of patents. To turn these research aims into testable hypotheses, the following section introduces the decision making logic of the EPO. We also introduce our set of explanatory variables (bibliographic indicators), which we use in our later analyses. Section

---

[7] As is well known to researchers in the field, bibliographic indicators are noisy measures of a patent's technological quality, and only a part of their variance captures quality-related phenomena.

[8] See for example U.S. Department of Commerce published public report PTD-9977-7-0001 entitled "Patent Quality Controls are Inadequate", attesting an increase in the USPTO's official error rate of more than 1000% from 1992 until 1996, as measured by assessment inconsistencies for a random set of cases. Initiatives to ensure "patent quality" by introducing quality control mechanisms (so called "second-pair-of-eyes" checks) demonstrate the concerns and uncertainties on the side of the offices regarding their current service quality level (SUEPO, 2002). *Nota bene* that patent office definitions of service quality are only concerned with the consistency/reliability of their service and less so with the absolute correctness of their judgment.

[9] We elaborate on the shortcomings of this approach in our discussion.

3.2 presents the fundamental logic of our estimations and Section 3.3 describes the independent variables. In Section 3.4, we present our testable hypotheses. Section 3.5 finally elaborates on the econometric specificities of the non-standard estimators chosen for the analysis.

3.2 Modeling the Decision Making Process of the EPO

EPO patent data is suitable for the study of the aforementioned research gap for various reasons. The paramount argument is that it allows us to simultaneously observe two procedural stages relevant for our analysis; namely, the granting procedure and the opposition procedure (including the ruling on the opposition). Essentially, as we elaborate upon below, mapping the EPO's decision on granting and opposition outcomes econometrically allows testing of whether an EPO patent – once granted – will survive any opposition. In turn, this will provide a good indication of whether EPO decisions reflect a normative understanding of 'quality' patents (Thomas, 2002).

Insert Figure 1 about here

Figure 1 reflects the stages of a European patent application. Upon application and examination request, the EPO makes decisions about the patentability of an application. Examiners assess applications in terms of several patentability requirements including novelty, inventive step (pendant in the US: non-obviousness), disclosure, and susceptibility to industrial commercialization. An application is granted patent status if it fulfills these requirements (Stage 1).[10] A granted patent can be challenged (i.e., attacked centrally for all designated states) within nine months, through an opposition procedure (Stage 2). For our purposes, it seems sufficiently correct to state that oppositions resemble (first instance) validity suits of a patent at the European level. Oppositions must relate to a patent's inability to meet the aforementioned requirements. The EPO decides on the opposition and one of three outcomes occurs: a patent is revoked, amended, or the opposition is rejected (Stage 3).

If patent offices acted in fully consistent ways (i.e. maximizing assessment quality regardless of its desirability) and the information on which granting decisions are based did *not* change from the date of grant until the date of opposition, the rejection rate of oppositions should be close to 100% (with a residual error arising from the "human factor"). This is because the general criteria for granting a patent initially and upholding a patent during opposition are virtually identical. Across all industries,

however, the rejection-of-opposition rate is far from 100%. We propose three explanations: either (A) a patent office acts inconsistently (according to our aforementioned definition) because it does not correctly assess information available at the time of granting about an invention's technological quality; (B) available information about a patent's underlying technology between the day of grant and the day of the opposition is different; or (C) both.

Each explanation assumes the existence of γ errors (i.e. falsely granted patents), although the γ error may or may not be systematic. Systematic γ errors (explanations B and C) shed a different light on the quality of a patent office's services than unsystematic γ errors (explanation A): e.g. the new information, originally provided during the opposition phase, may alter the assessment between patent grant and patent opposition. This may indicate that resource constraints, associated with generating prior art during the initial examination, drive the result. Alternatively, patent offices may assess identical information differently from granting to opposition phases: this is a larger subjective error than an error attributable to human factors[11]. In turn, assessments may appear erratic in this case.

3.3 Measuring (In)Consistency With Indicators of Technological Quality

To measure patent assessment quality – and decide for one of the potential explanations A through C – we can relate granting and opposition outcome decisions by the EPO to information relating to a patent's quality. This information, available to examiners and members of the opposition division, refers to a patent's novelty, inventive step (non-obviousness), disclosure, and susceptibility to industrial commercialization. In practice, patent offices examine information that is idiosyncratic and *very* complex (as complex as application and prior art documents, sometimes more than 100 pages of full text). For researchers, however, it is often not feasible to engage in activities of reviewing this information when examining a large sample of patents as we do. Instead, in large-scale patent econometric studies, typically researchers use a set of standard bibliographic indicators to capture phenomena relating to patent quality. Consequentially, in our study, we model a patent's (latent)

[10] Applicants may appeal against the denial of a patent grant, however, this is not the focal theme of this Paper (as mentioned before). See also Section 5.1.

[11] In our discussion we elaborate on the possibility that this human error is a matter of fact, in reality also reflects the resource constraints examiners face – namely a time constraint when evaluating the information. In that case, the human error would have systematic character, too.

objective technological quality with a set of bibliographic indicators, which have been used in prior studies. Table 1 provides an overview of our explanatory variables.

Insert Table 1 about here

Several of our proxies were originally validated as indicators of a patent's *economic/financial* value (see Reitzig, 2004 for a survey), which is admittedly not quite the same as being indicators of a patent's *objective technological quality*, despite the two being structurally related (see Reitzig, 2005). We list the rationales that link these proxy variables to objective technological quality in column 4 of Table 1. Although our indicators are blurred measures of a patent's technological quality, these should still capture a good part of its variance. In particular, works by Guellec and van Pottelsberghe (2000) and Reitzig (2005) support this allegation. In their work, the authors validate a large part of the aforementioned indicators as correlates of patent granting as well as opposition outcome decisions. We will bear the weaknesses of our approach in mind when qualifying our results. We discuss our indicators' limited ability to capture informational changes about patent validity over time. Finally, we also consider issues relating to an omitted variable bias (see 3.4 for more details).

3.4 Testable Hypotheses

Using the decision making logic of the EPO to shed light on our principal research questions, we formulate three complementary testable hypotheses to study within a framework of assumptions.[12] In testing our hypotheses relating to patent assessment consistency (Hypotheses 2 and 3), we assume that the EPO issues patents that are of a high quality along all dimensions and it does so – as best it can – in a constant regulatory environment (*Assumption 1*). We are *not* stating that such behavior is optimal: the assumption merely serves to set up our empirical test. We base our testing on the premise that we can proxy parts of a patent's technological merit by using bibliographic indicators (see above). We will test this premise separately for our sample (H1) and critically review this in our discussion section. In addition, we emphasize that we test relational and not causal hypotheses. Although we test whether our indicators are valid *correlates* of the information available to patent offices, we do not test whether the *indicators are ever causal* for the office's decisions.[13] We are aware that examiners and

[12] We relax some of these assumptions in Section 4.4 to shed more light on our theory-driven findings.

[13] We thank one of our referees for pointing us to the so-called "cookie-cutter" approach – a term coined in connection with US patent examinations. This term describes a situation in which examiners *do* base their

opposition divisions draw from different informational sources; however, we assume that our proxies are valid.

H1 speaks to our indicators' ability to capture technological quality. In line with earlier works (Guellec and van Pottelsberghe, 2000) we propose for our sample:

> *H1: For our given sample, the EPO's assessments of a patent's technological quality during validity decisions can be modeled by patent indicators, including backward references; family size; the PCT indicator; the number of applicants and inventors; and, the forward citations received until the date of grant.*

Essentially, if H1 cannot be rejected, we are confident that our indicators are proper measures of patent validity.

H2 then is the first comparative interpretation of the observable outcomes for patent grant and opposition outcome. To test H2 we use the observation that patent examiners (deciding whether to granting patents) and opposition divisions (deciding about opposition cases) share *partly* identical information. This time-invariant information is reflected in the time-invariant patent indicators (backward references; family size; PCT indicators; number of applicants and inventors) as well as those forward citations received until the date of grant.[14] If the EPO was able to accurately assess technological quality from the outset, this information would correlate similarly with their rulings in Stages 1 (granting) and Stage 3 (opposition outcome). Consequently, we propose:

H2: When the EPO's assessment decisions on a patent's technological quality are modeled by patent indicators (including, backward references; family size; PCT indicators; number of applicants and inventors; forward citations received until the date of grant) then there is no significant difference in the role of these predictors in granting and opposition outcome decisions.

---

decision directly on first-sight information of the type we use in our statistical analysis to model validity-related information. If this approach *had* been adopted during the grant of patens described in our sample as well, then the indicators we employ might in fact even be causal for the decision-making we observe. We do, however, have no information that EPO examiners ever adopted this cookie-cutter approach and hence assume that the indicators are mere correlates of the examiners' decisions.

[14] Note: as we explain in more detail in Section 4, we draw from the official European Patent Register EPOLINE in its version of December 2003. Based on all our inquiries – including talks with fellow researchers using EPO data as well as data experts at the EPO, the indicators we consider time-invariant should not change post-grant in the electronic registers.

Essentially, if H2 is rejected, patent assessment quality would, to an extent that it is measurable with bibliographic indicators (see before), be erratic in terms of reliability. However, even if the individual variables in H2 show similar roles in predicting granting and opposition outcomes, *overall* inconsistency may exist in the model and this result could be driven by unobservable heterogeneity. In other words, there would be a γ error (i.e., classifying an invention as patentable when objectively it fails to meet a quality threshold) in patent office's assessment, but we do not know what exactly drives it. Information that is subjectively new to a patent office[15] may be introduced during the opposition proceedings. This may change the γ type error, if there is one, from unsystematic to systematic (see above). Optimally we would be able to characterize informational change after grant when correlating them with the concordance of the EPO's decision making. Unfortunately, when using electronic sources it is difficult (not to say impossible) to capture the precise informational change per patent.[16] For our data, we find one indicator only that can capture *some* of the dynamics (information increase over time). Namely, this is the forward citation indicator that we assume is to some extent a measure of informational change over time (*Assumption 2*). Forward citations received after the date of grant, but before the end of the opposition procedure, should correlate with observable inconsistencies of a patent office. Consequently, with the aforementioned set of assumptions, we propose:

H3: If there is a significant difference in the EPO's rulings for granting and opposition outcome decisions of patents, then this difference should be correlated with a patents' forward citations received after the date of grant.

In summary, H1 proposes that patent offices systematically assess patents. This is proposed to occur when a patent is initially assessed (Stage 1) and when it is re-assessed following a challenge (Stage 3). H2 proposes that *systematic* consistency across stages should exist in these assessments. We propose patent offices use available information to assess technological quality to their best abilities. This implies that quality-related information available at the date of grant is assessed identically

[15] Of course it must not be objectively new since prior art is only judged until the date of filing.

[16] Currently, this information is only available in the form of scanned paper files that are substantially difficult to read/comprehend for the non-legal expert. Preparing this information for analysis is a challenging research project in itself that is left to future studies. Hence, we have no exact estimates of the precise informational change per patent. According to our interview partner in the Biotechnology Opposition Division of the European Patent Office, however, in most opposition proceedings additional information will be revealed

during examination (Stage 1) and opposition (Stage 3). We examine H3 only if inter-stage consistency does not hold. H3 involves testing our proposal that inconsistency can be explained and is not random; particularly, changes in information availability between grant and opposition are modeled. Failing this, we propose to examine an open-ended set of possible sources of inconsistencies.

At this point, a disclaimer appears in order. In essence, our hypotheses suggest testing a model of patent quality as being a function of certain observable components (bibliographic indicators). Of course, the idiosyncratic nature of each patent remains unobservable. To that extent, the model must be somewhat mis-specified because a number of unobservable components are excluded. This however, should be of little concern, unless there is reason to believe that the omitted variables affect the coefficient estimates of our independent variables (bias). For this to happen, the omitted variables would need to systematically – not randomly – relate to a patent's validity (the dependent variable) and correlate with (at least some of) the existing variables. While we can, by definition, never claim that we do not omit such variables, we deem the probability relatively low.[17] With this assumption (*Assumption 3*), however, our tests relating to inter-stage inconsistency are valid.

3.5 Methodological Aspects – Understanding the Scope and Usefulness of Variance Decomposition Discrete Choice Models for this Study

The empirical methods we introduce and use in this paper do not (yet) belong to the standard repertoire used by empirical researchers in the patent arena. An in-depth description of our methods and summary of the background literature we draw upon requires considerable space. For brevity, Appendix B describes our empirical approach in considerable detail. We dedicate this section (Section 3.5) to all other readers curious to obtain a fast working knowledge of the methods' scope and

---

that was not considered during the examination. Cases like EP 93 114 141, where the opposition division bases its judgment on the same set of information as a patent examiner, are considered to be the exception.

[17] We draw this inference based on two observations. First, from the set of established patent indicators (see Reitzig, 2004: 948, for an overview) we use the better part in this paper. The correlations (see Appendix A) among all our regressor variables are very moderate. Moreover, we do not find changes in the significance of individual coefficients when dropping variables from the models in Tables 3 to 5. Hence, we think that the likelihood of overlooking another important bibliographic indicator causing an omitted variable bias is low. Second, bibliographic indicators appear to capture parts of the variance of a patent's validity that are likely not proxied by other indicators we may be overlooking. See Reitzig (2004: 955) for an admittedly preliminary comparative test of bibliographic and other indicators.

comparative advantages relative to standard regressions used in the field, rather than familiarizing themselves with *all* the associated technical aspects.

Referring the reader to Figure 1, we can formulate our overall testing goals as follows: does our matrix of indicators X capture patent office decisions about patent validity in Stage 1 (grant) and in Stage 3 (challenge) (H1)? That is, are β_1 and β_3 significant? On average, are vectors β_1 and β_3 identical, recognizing that the difference between ε_1 and ε_3 must be accounted for (H2)? If β_1 and β_3 are identical the EPO interpreted information on technological quality (proxied by X) consistently during both granting and challenge phases. If β_1 and β_3 are not identical, which elements of the vectors account for any inconsistencies (H3)?

Testing H1 is a straightforward empirical problem for which we use standard binary discrete choice models (logit). Testing H2 and H3 is a bit trickier. In testing H2 and H3, we are hoping to make a direct comparison of how patent offices use available information to assess the quality of patents initially and upon challenge; this means comparing β_1 and β_3 (see Figure 1). The problem, however, is that in traditional choice models, estimates of β_1 and β_3 are inversely confounded with ε_1 and ε_3, respectively. Ultimately, a simple comparison of β_1 and β_3 is meaningless unless one can account for this confound. The random error components are different for numerous reasons; of major concern for this particular data set is that the set of Stage 3 observations are dependent on Stage 1 observations ("selection bias"). In particular, Stage 3 are the set of patents deemed to be of a high technological quality at Stage 1, but this has been questioned in the form of an opposition (the opponent implying that a patent was falsely granted). In turn, an ability to distinguish between valid and invalid patents in Stage 3, using the same set of indicators X as in Stage 1, is more difficult. That is, we *expect* the degree of error to differ in both stages, and that this will be reflected in ε_1 and ε_3.

One solution to our testing problem comes in the form of variance decomposition discrete choice models (VDDCM). These models make explicit use of an empirical feature in discrete choice estimations that standard methods tend to "neglect" for the sake of simplicity. Namely, this is the so-called distinction between "estimate" and "scale". Standard discrete choice models (both simple and nested ones) estimate a vector β that is, in reality, <u>not</u> β but β times λ. Here, λ is a scale parameter of the random component and commonly set to unity in estimations. λ is inversely proportional to the

variance component, ε. Variance decomposition discrete choice models do not ignore λ, - unlike most choice models (including random coefficient models) - but separate the β (true estimate) and the λ (confounding scale parameter) from one another.

By applying variance decomposition discrete choice models, we can disentangle spurious differences from true differences in assessments of technological quality during the grant and challenge phase. This allows us to correct for one of the most pressing problems created by the selection bias in the data for Stage 3.[18] Moreover, we can use the methods to determine the nature of potential assessment inconsistencies when testing H3.

4 EMPIRICAL RESULTS

4.1 Data – Stratification Criteria and Descriptive Statistics

For our analysis we chose biotechnology patent data from the 1970s and 1980s. This data allows us to examine a patent office's ability to assess the technological quality of patents consistently. At this time, the biotechnological industry was novel and emerging (Orsigeno 1989). In turn, it creates a "quasi-experimental" set-up that should capture the characteristics of similar cases concerning Merges (1999); namely, it is a situation where little prior art appeared in patent databases, but would be more likely to be documented in scientific publications and other sources. The examiners in our analysis faced similar informational challenges to those experienced by examiners today in fields such as software or nanotechnology.[19]

The selection of the industrial field was based on an updated version of the widely accepted OST INPI ISI classification by Schmoch (1994, personal note on an update from 1998). Biotech

[18] The selection bias is also likely to change the distribution in a different way which we do not explicitly account for but which we deem less important in the context of this paper. With Priest & Klein (1984), Cooter & Rubinfeld (1989), and Waldfogel (1991), we assume that an opponent's propensity to challenge a patent is driven by the value at stake as well as their subjectively perceived probability of winning the case. The opponents will initiate the opposition (see Lanjouw & Lerner, 1998; Harhoff & Reitzig, 2004; Reitzig, 2005) if: $\pi_{win} \cdot p_{opponent}(win \mid opposition) > \pi_{no\ opposition} - \pi_{lose} \cdot p_{opponent}(lose \mid opposition)$ where profits are those of the opponent. Hence, from a theoretical perspective we would expect not to see those patents in Stage 3 that are of very low techno-economic quality (since challenging them is worthless for the opponent, no matter how high the likelihood of winning) as well as those patents that are of extremely high legal quality (since challenging them appears pointless, no matter how valuable they are). Overall, however, we consider these two types of patents to be exceptions, delineating the margins of our overall distributions, and hence the issue negligible in this first study of patent assessment quality we conduct. This being said, it may be a worthwhile challenge to refine the estimation approach and custom-tailor it to the specific estimation problem in future studies on the subject.

patents were identified as showing one of the following IPC subclasses as their main classification: C07G; C12 M, N, P, Q, R, S. As of December 2003 (the date of data extraction) the European patent register contained 36,452 applications and 9,960 granted patents in these areas. 808 (8.11%) patents in the sample were opposed. For 558 of these opposed patents, a decision by the opposition division was observable. For the remaining part of patents, no clear ruling by the opposition division could be identified at that date (pending case either in opposition or appeal). Figure 2 shows the share of unidentified oppositions among the total sample versus the year of patent priority.

Insert Figure 2 about here

Figure 2 shows a peak of unidentified opposition outcomes in the year 1985, which translates into an ascent from 1988 onwards. Pending a better explanation, we attribute this increase to the share of opposition cases still to be decided in December 2003 by the opposition division of the EPO. As this paper focuses on decisions of the opposition division, we remove patents applied for from 1988 onwards. In doing so, we obtain a residual 7.2% of unidentified first ruling opposition cases (including appeals[20] in some cases) between 1978 and 1987.[21] Thus, the final data for analysis comprises 5,051 patent applications, out of which 3,162 were granted. A total of 334 granted patents in that period were opposed. The number of outliers dropped for the estimations is 24 cases (< 0.5%). Table 2 contains the descriptive statistics for the sample.

Insert Table 2 about here

The most important findings are the following. The rate of opposition in the biotechnology industry is high (10.56%), but it is lower than more litigious industries such as polymers (approx. 12% opposition between 1978 and 1990). Third parties invalidated patents in about 34% of all oppositions.

---

[19] Or, as some of our interviewees at the EPO puts it: "in the early 1980s, examiners in The Hague did not even have data bases where they could look up prior art in the area of biotechnology".

[20] Note: legally speaking, the opposition procedure *comprises* the (potential) subsequent appeal to the opposition. Appeals against the decisions of the Opposition Division can be made by all parties, patent holders and opponents, and are decided by the Board of Appeals. As is intuitively understandable, appeals to oppositions delay the final outcomes of opposition cases even longer. Hence, incorporating appeal decisions into opposition rulings would lead to even more data truncation problems than already present. Thus, for the purpose of this paper we *focus* on the *first decision* of the Opposition Division and not on the *final decision* by the Board of Appeals. We only include appeal decisions if they were decided before 2003 (extraction date) anyhow in order to avoid further truncation problems in the data. Admittedly, the data "quality" of our opposition rulings therefore differs as we incorporate the decision by the Board of Appeals in some cases. Even though we eventually deem this problem to be minor, this imperfection shall not be hidden from the reader.

Patents were amended in 32% of cases; opposition was rejected in 21% of cases because a challenge was not substantiated. In 13% of the remaining oppositions, the procedure was closed (7%) or no outcome identifiable (7%). Most of the explanatory variables appear within the "normal" range comparative to earlier studies, with some biotechnology specificities observable. On average, 2.9 references to patents of prior art were made by the EPO examiners during the European search procedure. This figure is slightly low (e.g., polymer patents average 3.5). Each patent cites 2.6 non-patent literature references as relevant state of the art. This figure is clearly higher than in more mature industries than biotech was in the 1980s. For example, polymer patents with priority dates 1977 and 1990 filed at the EPO quote only 0.5 non-patent references as relevant state of the art. Hence, as expected, the sample shows some of the experimentally desired properties (see above) relating to prior art. On average, patents were applied for in 9 states (i.e. countries signing the EPC) and almost three inventors (2.8) were involved in each application. The average accelerated examination requests is fairly low, with about one percent of all patent applications following the *Programme for Accelerated Prosecution of European Patent Applications (PACE).* 10% of filings were made according to Chapter II of the Patent Corporation Treaty (PCT); this is lower than for the entire population of EP patents, but higher than in polymers for the same time period. This indicates that applicants delay costly decisions in more than 10% of the applications by choosing the PCT II route. Finally, we computed forward citations for two different periods of time (namely for five-year and ten-year time windows after the application's publication date). We calculate different measures because of Hypothesis 3. In order to capture alterations in the information status about a patent's technological quality over time we use forward citations as a proxy. Optimally, we would like to distinguish which citations the published patent application received before grant and that received afterwards. By calculating forward citations for different time spans (as described above) we obtain a proxy for this distinction. We find an average grant lag (time span from the filing date until the granting date) of 5.4 years and an opposition outcome lag (time span from granting date until date of opposition outcome) of 5.1 years. Since forward citations were computed from the application date of a patent, 5-year forward

[21] The resulting "imperfection" of the data set appears acceptable, considering that with 7.2% unidentified cases (including oppositions that we "deemed to be withdrawn"), the "disturbance" of the opposition outcome variable is negligible.

citations capture information that was available for patent examiners during the first 5 years after application. Hence, for about a half of our patents, *all* information contained in this variable was revealed during the EPO granting phase. Ceteris paribus, the forward citations calculated for the 10-year time frame capture information revealed during the first 10 years after patent filing. For more than 95% of patents, this information falls at least partly into the period after patent grant. Finally, the difference between the two citation variables is an admittedly imperfect, but is a reasonable proxy for information revealed largely (i.e. for a large part of a patents in our sample) between grant and opposition outcomes.[22] As expected, the average number of forward citations in subsequent EPO search procedures increases with the length of the time window. For a five and ten year period it is 1.51 and 2.00, respectively.

4.2 Hypothesis 1: Systematic Assessment of Patents (Application and Challenge Stages)

According to Hypothesis 1, patent office assessments systematically relate to an observable set of quality indicators. These systematic assessments should be evident in the outcomes of Stage 1 (grant; not grant) and, if applicable, the outcomes of Stage 3 (patent revoked/amended; patent maintained/granted). We estimated a binary choice model for each stage. The systematic component of assessments in Stage 1 and Stage 3 are captured by β_1 and β_3, respectively. In Table 3, the binary choice models (Grant; Oppo) map the outcomes in Stage 1 (granted; not granted) and, if applicable, in Stage 3 (patent revoked or patent amended vs. patent maintained as granted).[23] The EPO's decisions were modeled using the aforementioned patent indicators (number of references to patent and non-patent literature; number of applicants and inventors; accelerated examination request; PCTI/II indicators; forward citations). We logarithmically transformed some of the explanatory variables when

[22] Since grant and opposition outcome lags vary, fixed citation time windows will never exactly capture pre- and post grant information for all patents. We picked the time windows in such form, however, that this measurement error should be minimal on average.

[23] Note: the coding of the binary outcome for Stage 3 is based on the following consideration. While there is only one technological quality threshold in Stage 1 (non-Grant vs. Grant) there are in fact 2 thresholds in Stage 3 (patent revocation vs. patent amendment and patent amendment vs. patent maintenance). Since, for econometric reasons, we need to boil down the complexity of Stage 3 to a binary decision we focused on the threshold in Stage 3 that would be most comparable to the granting threshold in Stage 1. Namely, whether a patent was upheld in exactly the same form as it was granted. Hence, we pooled the revocation and amendment of a patent on the one hand and the rejection of the opposition and the closure of the opposition procedure on the other (we dropped the cases of undecided oppositions). We preliminarily tested how our results would change if we coded the outcome in Stage 3 differently. These preliminary analyses suggest that the differences would not be radical, however, we did not inquire this in more detail.

(1) marginal effects of the variables on technological quality are decreasing (see Harhoff & Reitzig, 2004) and (2) the distributions of the individual variables were highly left-skewed (see Table 1).

We first estimate both models separately. The scale parameters (which inversely relate to the variance of random/error component) associated with each decision stage, λ_1 and λ_3, are not identifiable and arbitrarily set to unity. The estimates are provided in Table 3.

Insert Table 3 about here

We first examine the initial patent assessment decision in Stage 1 (model Grant), consistent with our framework of assumptions and in keeping with prior literature – in particular Guellec and van Pottelsberghe (2000). Our results suggest that we can proxy patent office assessments of technological quality using our set of indicators X – this supports Hypothesis 1. The model is overall well specified ($p<0.001$). We obtain individually and jointly significant coefficients for the variables relating to backward references to patent literature, accelerated examination requests, number of inventors, and forward citations received within five years of the publication date. Counter intuitively, the family size variable (coded as the ln[1+number of designated states]) correlates negatively with the likelihood of a patent being granted. We can only speculate about this finding; one plausible explanation is that large and cost-insensitive firms, with patenting tactics that cover wider product markets significantly more often, file patents for incremental inventions with a higher likelihood of not being granted. Admittedly, this explanation may be challenged.

The model of opposition outcomes in Stage 3 (model Oppo) suggests that the factors that were significant for granting initially are no longer significant. The overall model is badly specified (with $p=0.30$ the model is overall insignificant) – suggesting that the coefficients of the variables are not significantly different from zero; one important consequence would be to reject Hypothesis 1. For various reasons, we do not lean towards this interpretation. The most important one is sample size. In fact, when running *Model Grant* (Stage 1) on randomly chosen subsets of data comprising roughly 300 observations, we do not obtain significant results, either – whereas we do obtain them for N=5,027 observations (Table 3). Thus, the insignificant coefficients from Model Oppo (Stage 3) are likely to be

affected by the comparatively small N – and we estimate inefficiently.[24] Given this, and considering the substantial empirical evidence of our indicators' ability to capture phenomena related to technological quality in Stage 1 (see Table 1, see Section 3.3), we are inclined to give more weight to the findings of that model. In turn we proceed with our analysis under the assumption that H1 is fulfilled for our specific set of data.

4.3 Hypothesis 2: Inter-stage Consistency in Assessment of Patent Quality

According to Hypothesis 2, patent offices use quality-related information in a systematic and *identical* fashion to assess the underlying quality of a patent initially (granting phase) and upon reassessment (challenge phase). Our time-invariant indicators capture information that should be identical at the day of grant and at the day of opposition. Comparing the estimates relating to these indicators at each stage will reveal whether the EPO is consistent in their decision-making. We do this in model 4 but account for potentially spurious results in the β's arising from differing error structures (as reflected in the ε's and, hence, λ's, see Section 3.5). Econometrically speaking, in this model we test whether coefficients for X are consistent when variance components from the two data sets on grant (stage 1) and opposition (stage 3) are allowed to differ. That is, we introduce the restriction $\beta_1 = \beta_3 = \beta$, given $\lambda_1 \neq \lambda_3$. As described in detail in Appendix B, the data for stages 1 and 3 are pooled for this test and treated as independent data sets.[25] We interpret the overall goodness of the model by focusing on the likelihood ratio values.

Insert Table 4 about here

The estimates are shown in Table 4. It is clear that the successful rescaling procedure is questionable, given a negative scale ratio. Specifically, since the scale is inversely related to the variance, in a correct and acceptable model this ratio must be positive. This is confirmed when comparing the proposed model to the unrestricted model. The likelihood-ratio value is 23.99, which is greater than the χ^2 value of 18.31. As a result, we conclude that once differences in variability are

[24] This problem of inefficient estimation should become smaller once we pool the sub samples (see Tables 4,5, and 6) in order to test hypotheses 2 and 3. Note again that a simple comparison of the absolute sizes of the coefficients for X in Models Grant (Stage 1) and Oppo (Stage 3) to test H2 and H3 are meaningless since they may be the result of a spurious scale artifact (biased).

[25] See FN 19: resulting imperfections from this assumption in our empirical design offer, in our eyes, scope for future research, however, they do, in our understanding, not preclude our first careful analysis presented in this paper.

accounted for, the model 4 in which we propose that the office shows consistent assessments about a patent's technological quality across both granting and opposition stages (Hypothesis 2) is not supported (its H_0 is not rejected). In turn, one should examine the restricted estimates with caution[26].

4.4 Hypothesis 3: Theoretical Sources of Inter-stage Inconsistency in Assessment of Patent Quality – Informational changes between grant and opposition

According to Hypothesis 3, potential inconsistencies are attributable to incremental knowledge obtained after the day of grant. That is, we attempt to account for beneficial hindsight knowledge provided to opposition division members. Using incremental forward citations received largely after grant as a proxy for hindsight information, model 5 tests H3 in a way that is comparable to our testing of H2, but for one important difference: in model 5 we estimate a pooled model of the two assessment stages where some *combination* of assessment homogeneity and assessment heterogeneity by the EPO is imposed, while still allowing variance heterogeneity to exist. It is important to note that we restrict the model so that the assessment heterogeneity may *only* be explained by the incremental forward citations, our proxy for hindsight information. This is in line with our theoretical expectation (H3). Of course, patent offices should systematically consider *other* (= time *in*variant) aspects of technological quality in the same way (i.e., consistent) across the two periods of assessment.[27] Econometrically speaking, in model 5 we introduce identical restrictions to model 4 ($\beta_1=\beta_3$) and the same relaxed assumptions relating to the error terms ($\lambda_1\neq\lambda_2$), but introduce the β representing information changes (k=10) as a heterogeneous parameter (i.e., $\beta^{k=10}_{\_Stage1}\neq\beta^{k=10}_{\_Stage3}$).

Insert Table 5 about here

The results indicate that Model 5 proposes a set of restrictions that do not enable the model estimates to converge to an acceptable solution, That is, the data sets cannot be combined simply by

[26] We caution the reader to interpret the model estimates. Essentially, we proposed that we can estimate a model in which the bibliographical indicators of quality were used in an identical fashion to judge quality in stages 1 and 3, recognizing the errors at the two stages differ. Hence, this would imply only a *single* set of estimates are required. This model was rejected, which indicates that the 'k' estimates themselves are misleading.

[27] The reader may note that in model 5 we drop the forward citations received until grant (5-year frame) from the set of homogenous parameters. We do so to avoid obtaining potentially confounding collinearity effects between the forward citation measures. Essentially, our parameterization in model 5 therefore gives the data the "maximum chance" to show consistency across Stages 1 and 3.

rescaling and with a single heterogeneous parameter. This is since the model introduced is overly restrictive (note: this result holds even when introducing a heterogeneous intercept term to allow for different propensities in the granting or upholding of patents by the office; the results of this test are not shown, but available upon request). The results suggest that a model of consistency in patent assessment (as proxied by indicators of patent quality) over assessment stages cannot be supported, even when allowing for different levels of error and possible information changes as proxied by incremental forward citations.

We can, therefore, formulate an important interim summary: within the framework of our assumptions, particularly with an assumption that bibliographic indicators can act as proxies for validity-related information, we do not find that the EPO assessed technology related information consistently between grant and opposition during the 1980s in the area of biotechnology. The inconsistency we observe is not solely attributable to informational change between grant or challenge decisions.

While we think that our tests support our conclusions quite substantially, we reiterate that it is not the aim of the paper to downplay any individual's efforts at the EPO to deliver the best possible service during the period we study. We do not make conclusions of low assessment quality based on an isolated empirical affirmation. Instead, we prefer to examine (self-)critically which technical premises and theoretical aspects of our research design should be relaxed to explain our empirical findings comprehensively. The following Section is dedicated to doing this.

4.5 Investigating Further Potential Sources of Inter-Stage Inconsistency

In testing our research hypotheses, we concluded that the office may use indicators of patent quality in a systematic fashion (H1), but do so in a inconsistent fashion at the two stages of assessment; namely upon initial patent assessment and upon challenge. This was evident even when correcting for the confound in parameter estimates introduced by difference in error structures (H2) and allowing for informational changes to be heterogeneously assessed (H3). In other words, hypotheses 2 and 3 are rejected.

To this extent, we are motivated to investigate *further* reasons why our proposal of homogeneity (and hence, consistency), as well as theoretically comprehensible heterogeneity, is

rejected. Could it be that simply one or a few bibliographical indicators (i.e., parameters 2 through 9) are the pieces of information which reflect inconsistent assessments? Or could it be that a completely different set of criteria is used to judge patent quality in challenge stages and that the rectification of γ errors by the office is difficult to comprehend relative to their assessments made in Stage 1? Perhaps the answer is somewhere in between? In models 6a and 6b we propose and test whether inconsistency in the EPO's ruling – as found in 4.3 – is driven by only a few of the informational indicators we use; thereby this test relaxes the assumption that hindsight knowledge may only be captured by incremental forward citations (*Assumption 2*). Secondly, we relax our framework by dropping the assumption which states the EPO worked with constant conditions over time (*Assumption 1*); in other words, we allow for the γ errors to be driven by pure time-trend contingencies. These contingencies may be exogenous to examiners and members of opposition divisions, or reflect their own behavior (e.g. their learning about the new technologies).

To relax Assumption 2, we estimate a model that is similar to model 5, however, we make no prediction or impose no restriction on assessment heterogeneity. Any of our indicators may now capture heterogeneous assessments. In searching for sources of heterogeneity (inconsistency), we tested several models to identify a potential set of parameters accounting for the heterogeneity. Table 6 presents the results of two specifications which emerged as the most significant and stable ones after comprehensive testing. Moreover, we split Table 6 into columns A (restricted parameter set without incremental forward citations) and B (parameter set including the incremental forward citations received after granting). By contrast, the two models provide an additional indication on the usefulness of our incremental forward citations to measure informational change after grant. It also serves as a robustness check for our rejection of H3.

Insert Table 6 about here

The findings show distinct variables are driving the variance heterogeneity across assessment Stages 1 and 3 in both models 6a and 6b. Namely, these are accelerated examination request; number of inventors; absolute number of forward cites received within 5 years after publication (in model 6a); and incremental forward citations received between 5 and 10 years after application (in model 6b). We cannot offer a rational explanation for why a patent office would interpret information ***correlated*** with

acceleration requests or the number of inventors differently across the granting and opposition stage. On the other hand, it is intuitive why the incremental forward cites (received between five and ten years after publication) exhibit variance heterogeneity. Specifically, the variable was included to capture potential informational changes over time in model 5. It appears to show the desired effect only in a model which is less restricted (6b). It is somewhat puzzling that forward citations within the 5-year period are also driving heterogeneity (model 6a). Theoretically, this variable should not exhibit any variance heterogeneity across assessment stages. The correlation structure among the independent variables suggests, however, that a moderate correlation between incremental forward cites and 5-year time frame forward cites is driving this result. This correlation as well as the measurement error of this variable in picking up pre- *vs* post-grant information leads us to revisit the ability of our measures to pick information status at different stages, and address this in our discussion.

To relax Assumption 1, we re-ran models 6a and 6b including a set of time dummies to capture the variance in the EPO's decision-making on patent validity over time above and beyond its (re-)assessment of technological quality for identical patent (application)s. We created time dummies for each patent filing year to capture the contingencies of providing services that are (a) exogenous from an examiners perspective (i.e. change in fee structures, change in number of applications, etc.), and (b) reflect examiner behavior (e.g. learning).[28] Our results are summarized in Figure 3:

Insert Figure 3 about here

The reader may note the time trend created by the significant coefficients of the time dummies on the probability of a patent being held valid (all else being equal). The years 1978 and 1979 were combined due to a lack of observations in Stage 3 for patents being filed in these years, and acted as the reference year in the estimation (hence, the coefficient equal zero by default). Of importance, but not visible from Figure 3, is that almost all remaining indicators of technological assessment quality are insignificant in VDDC models 6a and 6b (except backward references to patent literature for both models and PCT I for Model 6a)[29] Finally, with the exception of the two (one) indicators, model 6a

[28] This approach does not allow us to disentangle fully exogenous effects from the perspective of patent examiners/opposition divisions from endogenous ones.

[29] Readers familiar with the use of European patent data may wonder whether these last results could, in part be driven by the fact that some of our measures, while time-*in*variant for the individual patent, show distributional changes over our sampling period (e.g. the maximum number of states to be designated changed from 1978 until 1987). We can not exclude that this is the case to some extent; however, *if* it was, this would

(6b) that include additional time dummies show overall support for H2. In summary, for our sample – and within our framework of assumptions – the EPO was consistent in its rulings on patent validity, but its decisions were not based on assessments of patents' technological quality; instead decisions were driven by changing environmental (internal and external) conditions!

5 DISCUSSION

5.1 Limitations

Our discussion begins with some considerations about the robustness of our findings, as well as a repeated disclaimer.

Conceptually, our analysis focuses on only one aspect of patent assessment quality, although it is an important one. We study assessment inconsistencies arising from errors (we term these γ errors), which occur when a poor patent – one that should never be granted – is granted. The isolated inspection of γ errors may be misleading when assessing overall/total patent assessment quality.[30] This is because, theoretically, the desirability of γ errors is dependent upon patent offices having mechanisms in place to correct for β errors (i.e., falsely rejected patent applications during grant). If an institution has a harsh granting procedure and the number of falsely rejected patent applications is high, then the risk for γ errors occurring is likely to be low; the need for institutional features to correct for γ errors (falsely granted patents), therefore, will also be low. Conversely, if there is a mechanism to correct for β type errors, there must be mechanisms in place to deal with γ errors. The EPO has mechanisms in place for correcting both β and γ type errors. β errors (falsely rejected applications) can be corrected through the "appeal to grant" procedure. A profound analysis of this institutional feature would be beyond the scope of this paper, but the reader may note the following descriptive statistics. Out of our sample of 5,051 *patent applications*, 81 grant examination decisions were appealed. Out of the 3,162 *granted patents*, 334 were opposed and 79% of those were admitted for opposition. We have no objective yardstick against which to measure these figures; we do, however,

---

render our results even more powerful – unless one assumes that bibliographic measures are unsuited to proxy technological quality.

[30] We thank one of our referees of this paper for encouraging us to delve deeper into this discussion – which we deem very important.

find the number of appeals to be relatively high considering that the discrimination between patent validity in Stage 1 should, on average, be easier than in Stage 3 (see our aforementioned arguments). Hence, we interpret this figure as an indication that the EPO's "appeal against examination" procedure is avidly used. Thus, we do further assume that the desirability for the EPO to commit γ errors is limited, reinforcing our argument to study γ errors as an important indicator of overall patent assessment quality.

Empirically speaking, we recall that our results are based on the assumption that the bibliographic indicators we adopt allow operationalizing a patent's technological quality. In general, we consider this assumption unproblematic since it is supported by an extant literature in the field. Moreover, model Grant (Table 3) indicates that this is a reasonable assumption for our specific data if N is large enough, too (see also Appendix Table A2). We are aware that our indicators are *correlates* of the EPO's informational decision making basis and not *causal* for EPO's decision-making. We do not claim that our indicators capture the technological quality concept: we are aware that our estimations may be systematically underspecified. However, we do not see why at least those indications about a patent's quality we can measure should exhibit different effects in patent quality assessments across different stages (granting vs. opposition procedure). Moreover, we assume that the incremental forward citations received between 5 and 10 years after patent grant are, for the majority of patents within this sample, a good first proxy for the informational change regarding a patent's technological quality between the day of grant and the day of the opposition outcome (*Assumption 1*). The average time lags for patent granting and opposition decisions as observed for our data render this assumption plausible. We know, however, that the variable dos not capture the entire change in information for some patents. Moreover, our incremental forward citation measure is unlikely to reflect oral information on "prior use" of a technology introduced during the opposition phase. In any case, the proxy does not allow us to determine the type of informational change with precision. Empirically speaking, the correlation between the different forward citation measures and their similar impact on consistency (models 6a and 6b) caution us not to over-interpret related results. Finally, except for our final estimations (see Figure 3), we assume that the conditions with which the EPO operates over time are constant and, if it is not constant, this should not affect the EPO's decision

making (*Assumption 2*). We will critically review the suitability of this assumption in the following section and discuss implications of our findings relating to assessment (in)consistency.

5.2 (In)Consistent Technology-Based Assessments?

Stimulated by the ongoing discussion about patent assessment quality, this paper sought to generate robust empirical evidence on various questions. First, we asked: what is the degree of technology-based decision-making consistency between validity-related decisions during patent grant and patent opposition? The question relates back to two fundamental issues of patent quality; namely, whether patent granting procedures are reliable by (a) categorizing patent applications along a yardstick of technological quality; and (b) ensuring that a patent survives a subsequent "validity suit" in emerging patenting areas. The answer to this question is easily summarized: within our framework of assumptions, Hypothesis 2 is rejected. This means that, for the entire sample of biotechnology patents applied for between 1978 and 1987 at the EPO, we do not observe consistent rulings by the office in the sense that the EPO assessed *identical* information on patentability requirements (technological quality) differently at the day of grant and the day of the opposition.

5.3 Sources of Inconsistency

Second, we asked: what are the sources of inconsistent judgments between patent grant and challenge (opposition)? While our analysis does not permit us to answer this question conclusively, we offer the following response: the technology-related informational change observed between the day of grant and the day of the opposition outcome is, to the extent that it is captured by the incremental forward citations, not entirely driving the assessment inconsistencies. This finding follows from our rejection of H3. Essentially, the interpretation of *identical* information, rather than the interpretation of *different* information, drives inconsistent assessments.

5.4 Desired Inconsistency?

Third, we posed the question: what we can infer from these findings on the level of service quality provided by the EPO in the area of biotechnology during the 1980s? While our assumptions and the limits of our empirical design require drawing conclusions carefully, the following appears noteworthy. Our results do not support the finding that patent validity-related decisions were based on consistent assessments of objective technological quality.

We *do not* find strong indications that the EPO attempted to *optimize* its service quality. To illustrate, consider the following counterfactual chain of thoughts: had the inconsistencies between patent grant and patent opposition been largely attributable to a *change in information about a patents' validity between the day of grant and the day of opposition*, then this would have been an indication that the EPO allocates only limited resources to the *search of prior art* to patent examiners. Consequently, this would have led to the discovery of information during opposition that was 'overlooked' during examination. If this was the case, however, one might suggest that the EPO provided service quality *conditional* upon resource constraints (see Merges, 1999), rather than being entirely "rationally ignorant" (see Lemley, 2001). Such a conditional resource allocation would suggest the EPO did optimize its service quality. Empirically, however, we do not obtain any findings suggesting information changes can explain inconsistencies of the EPO; in turn, we therefore second-guess that the EPO optimized its service quality. Admittedly, the limited explanatory power of our measure of informational change over time (incremental forward citations) asks one to treat this conclusion with caution. Whether our findings *truly* reflect an optimal allocation rationale for *search resources* between granting and opposition by the office remains a partly open question to be addressed in future research.

Quite clearly, however, inconsistent assessments regarding *identical* information during the grant and the challenge phase is not desirable in *any* circumstances.[31] Essentially, this inconsistency is a "human" error. If available, patent offices should consider options to reduce this error. For example, patent offices could grant examiners more time or offer greater training and education – especially in "emerging" patenting areas. Given Merges' suggestions, these options may benefit patent offices in the long-run as it requires considerable time to find and interpret prior art. The time trend we plot in Figure 3 may be attributable, not only to the changing environmental conditions with which examiners operate, but also to their "learning" in new technological areas.[32] Such measures, however, do not guarantee that "undesired" inconsistency will vanish. If there is a chance that patent offices are, for

[31] In fact, only with Lemley (2001) one might argue that it may be a reflection of an optimal resource allocation policy of a patent office if examiners and opposition division judge identical information differently. Namely, if one argues that not only *resources for the search of prior art* but also *resources for the interpretation of prior art* should be allocated mainly to the opposition division. We do not elaborate further on this thought for reasons given in 2.1.

[32] We thank one of our reviewers for sharing this thought with us.

one reason or another, incapable of guaranteeing a high level of consistency at reasonable costs – or normative legal quality (Thomas, 2002) – in emerging technologies, what would other options be to improve the system? Is there value to introducing a "grace period" for granting patents in new technologies? In other words, should society wait and observe a technological area *before* it acts and offers mechanisms for protection? What are the opportunity costs of this approach? These questions are highly relevant for future research endeavors examining patent quality.

6 CONCLUSIONS AND FUTURE RESEARCH

Stimulated by the ongoing discussion about patent quality, this paper sought to generate robust empirical evidence on the following two questions: (1) Did the European Patent Office (EPO) consistently base (repeated) patent validity decisions on its judgments of technological quality? (2) What were the sources of inconsistent judgments (if any) between patent grant and challenge? Moreover, we considered what our findings reveal about whether the EPO allocates resources optimally between the grant and the challenge phases.

Using data on European biotechnology patents filed between 1978 and 1987, we show that the EPO's decision making on a patent's technological quality during granting and opposition phases ("validity suit") was inconsistent; to the extent we can measure this using bibliographic indicators. Moreover, we do not find compelling evidence that the inconsistency indicates "optimal" resource allocation in a way that can be explained by informational increases on a patent's technological quality from granting until the end of the opposition procedure. While there is some indication that information about a patent after its granting accounts for some of the differences in patentability assessments over time, we have no conclusive empirical evidence for this. This implies that examiners and opposition divisions judge identical information in different ways. We argue this is undesirable. Our results are subject to several caveats. In particular, issues of specification (unobserved contingencies), specific model assumptions (correlation of error term structures), and preliminary character of our measure for informational change over time (= incremental forward citations) may distort our findings. In addition, we focus on one emerging patenting area, prohibiting us from making any undue generalizations.

This being said, we would prefer to think that we make several contributions and extend the current debate on patent quality. We believe that this paper provides the first large-scale empirical test of patent assessment quality according to a definition that captures both economic and legal dimensions of patent quality. In addition, our tests are carried out within only one office and one industry, but earlier studies compare different industries with one another and face potential problems relating to the neglect of important contingencies.

We know of no prior study that exploits patent data for the aforementioned research question beyond the level of rather simple comparisons of average indicator variables. Important information regarding the consistency of decision-making of a patent office may be hidden in (or distorted by) the variances of these indicator variables; our study is the first to exploit the heterogeneity and richness of patent data to shed more light on a patent quality discussion.

As in many research endeavors, our study has left us with many questions. Some issues we deemed important were raised in the discussion, and, in our view, present stimuli for future research. In order to understand the generality of our findings, a comparison of the biotechnology industry in the 1980s with a mature patenting area (e.g., modern polymers) may provide insights. Such a study comparable to this one, but using patent data from a different technology, could shed light on the question whether Merges' (1999) criticism is limited to new patenting areas, or whether problems may exist in other technological areas. In order to understand the sources of inconsistent rulings, qualitative research may be worthwhile. We see little space for "squeezing" existing bibliographic data further than in this context, and we would expect good survey or archival data to help identify these sources. This particularly applies to the variables capturing the precise informational change between patent grant and opposition. Finally, our study focused on γ errors (false grants) exclusively. A comparative study focusing on β errors (falsely non-granted applications) might reveal additional insights about the total service quality provided by the EPO.

Acknowledgements and disclaimer

We are grateful to Dr. Leonardo Galligani and Dr. Christopher Heath (both Board of Appeals, the European Patent Office) for sharing some of their knowledge on institutional legal details with us. Despite it being obvious, we reemphasize that the opinions expressed in this paper, however, are entirely our own and do not necessarily represent the official opinion of the EPO nor the personal opinions of those EPO employees who merely helped us understand their institution better. We also thank Stefan Wagner (University of Munich), Jordan Louviere, and Nada Wasi (University of Technology, Sydney) for thoughtful comments. As usual, Ralph Schories provided excellent research assistance when extracting the data and computing the indicators. Finally, Markus Reitzig acknowledges support (research time buyout and travel means) from the Danish Social Science Council under grant schemes RIPE (Network for the Research on Intellectual Property Economics) and LOK Biotechnology Business while he was working at Copenhagen Business School. The usual disclaimer applies, and the authors are responsible for all other errors to the extent that they have accrued from their work. This does not include potential shortcomings of the original EPO patent register raw data we draw from, which we may not have been able to eliminate fully despite all the consistency checks we used.

REFERENCES

Allison, J. R. & Tiller, E. H. (2003). The Business Method Patent Myth. *Berkeley Technology Law Journal, 18 (Fall).*

Barton, J. (2001). *Non-obviousness*. Manuscript drafted for the Franco-American Conference on Intellectual Property, Berkeley.

Ben-Akiva, Bradley, M.M., Morikawa, T., Benjamin, J., Novak, T., Oppewal, H. & Rao, V. (1994). Combining Revealed and Stated Preference Data. *Marketing Letters 5 (4),* 335-50.

Ben-Akiva, M. & Lerman, S. (1985). *Discrete choice analysis (sixth ed.)*. London, England: The Massachusetts Institute of Technology Press.

Carpenter, M., Cooper, M. & Narin, F. (1980). Linkage Between Basic Research Literature and Patents. *Research Management (March),* 30-35.

Cooter, R. & Rubinfeld, D. (1989). Economic Analysis of Legal Disputes and Their Resolution. *Journal of Economic Literature 27(September),* 1067-1097.

EPO (2003). *Annual Report*. The European Patent Office, Munich.

Ernst, H., Leptien, C. & Vitt, J. (2000). Inventors Are Not Alike: The Distribution of Patenting Output Among Industrial R&D Personnel. *IEEE Transactions of Engineering Management 47(2),* 184-199.

Gallini, N. (2002). The Economics of Patents: Lessons from Recent U.S. Patent Reform. *Journal of Economic Perspectives 16(2),* 131-154.

Graham, S.J.H., Hall, B.H. , Harhoff, D. & Mowery, D.C. (2002). Post-Issue Patent "Quality Control": *A Comparative Study of US Patent Re-examinations and European Oppositions*. NBER WP 8807.

Green, J. R., & Scotchmer, S. (1995). On the Division of Profit in Sequential Innovtion. *RAND Journal of Economics 26(1),* 20-33.

Guellec, D. & van Pottelsberghe de la Lotterie, B. (2000). Applications, Grants and the Value of Patent. *Economics Letters, 69(1),* 109-114.

Harhoff, D. & Reitzig, M. (2004). Determinants of Opposition against EPO Patent Grants: The Case of Pharmaceuticals and Biotechnology. *International Journal of Industrial Organization, 22 (4),* 443-480.

Hensher, D., Louviere, J., & J. Swait (1999). Combining Sources of Preference Data. *Journal of Econometrics 89,* 197-221.

Hoffman, Paul J. (1960). The Paramorphic Representation of Clinical Judgement. *Psychological Bulletin, 57,* 116-31.

Lanjouw, J. O., Pakes, A., & Putnam, J. (1998). How to Count Patents and Value Intellectual Property: Uses of Patent Renewal and Application Data. *The Journal of Industrial Economics XLVI (4)*, 405-33.

Lanjouw, J. O. & Lerner, J. (1998). The Enforcement of Intellectual Property Rights. *The Annales d'Economie et de Statistique, 49/50*, 223-246

Lemley, M. (2000). *Rational Ignorance at a patent Office*. Boalt School of Law, University of California at Berkeley Working Paper.

Louviere, J. (2001). What If Consumer Experiments Impact Variances as well as Means? Response Variability as a Behavioral Phenomenon. *Journal of Consumer Research 28 (2) December*, 506-11.

Louviere, J., Fox, M. & Moore, W. (1993). Cross-Task Validity Comparisons of Stated Preference Models. *Marketing Letters 4 (3)*, 205-13.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka, Ed, *Frontiers in Econometrics,*. New York: Academic Press.

Merges, R.P. (1988). Commercial Success and Patent Standards: Economic Perspectives on Innovation. *California Law Review 76(803)*, 805-876.

Merges, R. (1999). As many as Six Impossible Patents Before Breakfast: Property Rights for Business Concepts and Patent System Reform. *Berkeley Tech. L. J. 14*, 577-615

Narin, F., Noma, E. & Perry, R. (1987). Patents as Indicators of Corporate Technological Strength. *Research Policy 16*, 143-155.

Nordhaus, W. D., 1967, *The Optimal Life of a Patent*. New Haven.

Orsenigo, L.(1989). *The emergence of biotechnology. Institutions and markets in industrial innovation*. London: Pinter.

Parasuraman, A., L. Berry, and Zeithaml, V. (1991).Refinement and Reassessment of the SERVQUAL Scale. *Journal of Retailing 61 (4)*, 420-450.

Priest, G. & Klein, B. (1984). The Selection of Disputes for Litigation. *Journal of Legal Studies 8*, 1-56.

Quillen, C., & Webster, O. (2001). Continuing Patent Applications and Performance of the U.S. Patent Office, mimeo.

Reitzig, M. (2004). Improving Patent Valuation Methods for Management – Validating New Indicators by Analyzing Application Rationales. *Research Policy 33 (6/7)*, 939-957.

Reitzig, M. (2005). *On the Effectiveness of Novelty and Inventive Step as Patentability Requirements – Structural Empirical Evidence Using Patent Indicators*. SSRN WP 745568.

Sampat, B., Mowery, D. & Ziedonis, A. (2003). Changes in university patent quality after the Bayh–Dole act: a re-examination. *International Journal of Industrial Organization 21*, 1371-1390.

Samuelson, P. (2004). *Legally Speaking: Why Reform the U.S. Patent System?* Mimeo, University of California at Berkeley.

Sanyal, P. & Jaffe, A. (2005). *Peanut Butter Patents Versus the New Economy: Does the Increased Rate of Patenting Signal More Invention or Just Lower Standards?* WP, Brandeis University.

Schmoch, U. & Kirsch, N. (1994). *Analysis of International Patent Flows*. Karlsruhe, Fraunhofer Institut für Innovationsforschung and Systemtechnik Karlsruhe.

Scotchmer, S. & Green, J. (1990). Novelty and Disclosure in Patent Law. *RAND Journal of Economics 21(1),* 131-146.

SUEPO (2002). *A Quality Strategy for the EPO*. SUEPO WP, Munich.

Swait, J. & Louviere, J.J. (1993). The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models. *Journal of Marketing Research 30 (3),* 305-14.

Thomas, J. (2002). The Responsibility of the Rulemaker: Comparative Approaches to Patent Administration Reform. *Berkeley Tech. L.J. 17*, 728-761.

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

Waldfogel, J. (1998). Reconciling Asymmetric Information and Diverging Expectations Theories of Litigation. *Journal of Law and Economics 49 (2),* 451-476.

Figure 1

Decision Tree: From Application to Opposition Outcome

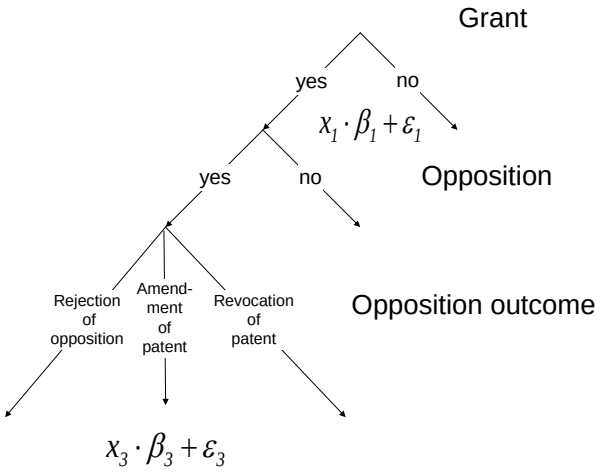


Figure 2

Share of Unidentified Opposition Outcomes among All Opposition Cases in Biotechnology vs. Year of Application

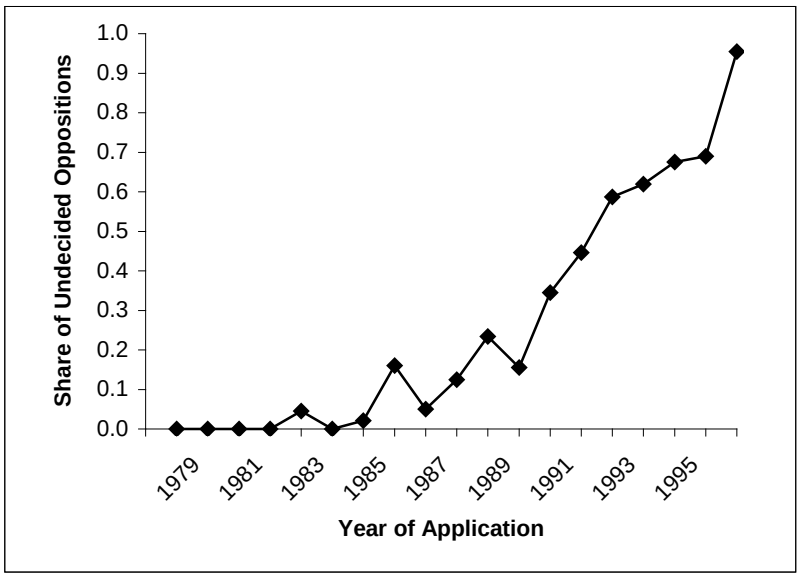
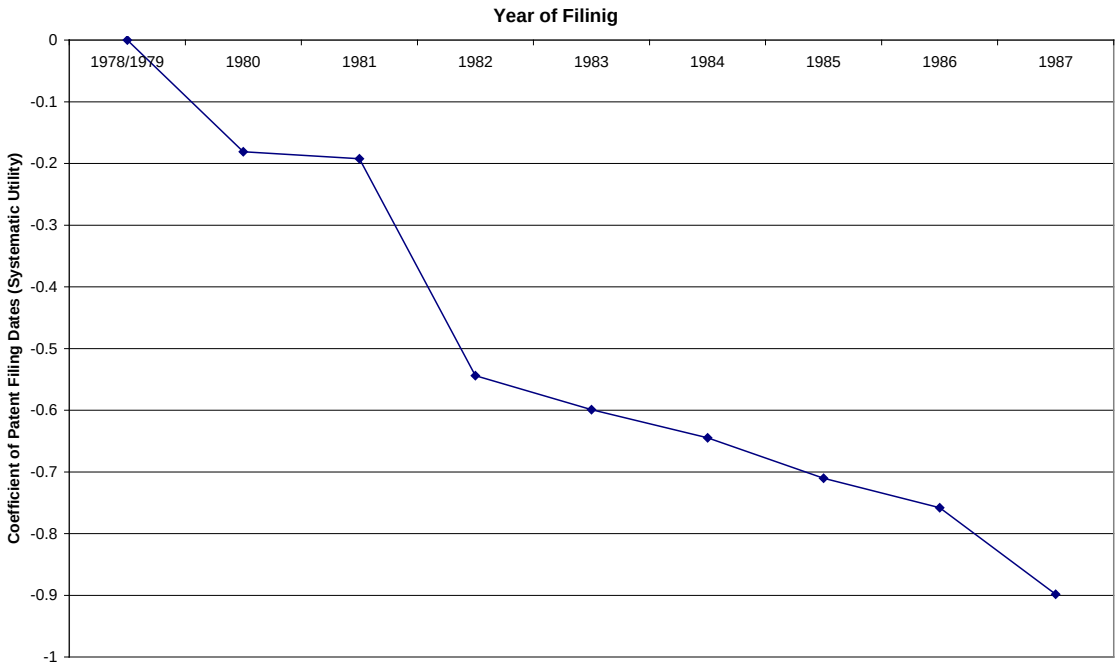


Figure 3

Effect of Application Year on (Pooled) Probability of Patent being Valid in Covariance Heterogeneity Model with Mixture of Heterogeneous and Homogeneous Systematic Assessment Parameters[33]



Box 1

Types of Error in Assessment of Patent Quality

Patent Office Assessment	True/Objective Patent Quality: Good	True/Objective Patent Quality: Poor
Good / Patent granted	α	γ
		Threshold A
		Threshold B
Poor / Patent rejected	β	α

[33] Note that in order to calculate the time intercepts in Figure 3 we had to adjust the parameterization of model 6a slightly in order to avoid multi-collinearity effects of the RHS variables.

Table 1

Patent-Based Measures and their Link to "Patent Validity" (Objective Technological Quality)

Variable	Detailed description of measure	Coding	Link to technological quality (degree of novelty and inventive step):rationales (-: moderate negative correlation; +: moderate positive; ++: strongly positive)	Basic reference
Number of Backward Citations to a patent Literature	Number of patent references to the state of the art that are actively quoted by a patent	Logarithmic	++: the more developed a technological area, the greater the involvement of professional (corporate) inventors, and hence the higher likelihood of the focal patent being of high technological quality -: the more developed a technological area, the more marginal the focal patent's contribution	Narin, Noma & Perry (1987)
Number of Backward Citations to the Non-Patent Literature	Number of non-patent references to the state of the art that are actively quoted by a patent	Logarithmic	+: the closer a patent application is to "basic" research, as reflected by the non-patent references, the higher its technological quality (relevant for scientific references) -: the less "scientific" a backward reference, the lower the technological quality (relevant for non-scientific references)	Carpenter, Cooper & Narin (1980)
Number of Designated States (Family Size)	Number of states	Logarithmic	+: the higher the applicant's willingness to pay for enlarged territorial protection, the higher a patent's value (and, hence, potentially its technological merit) -: the larger a potential market for a patent, the higher the likelihood of the focal patent being an incremental contribution and therefore of low technological quality	Lanjouw, Pakes, & Putnam (1998)
Number of Applicants	Number of applicants (natural and legal persons) involved in the application for a patent	Binary	++: the more applicants contribute resources to the research and development process underlying the focal patent, the higher the resulting technological quality	Guellec and van Pottelsberghe de la Potterie (2000)
Number of Inventors	Number of inventors (only natural persons) involved in the application for a patent	Logarithmic	++: the more inventors participate in the research and development process underlying the focal patent, the higher the resulting technological quality	Ernst, Leptien, & Witt (2000)
Number of Forward Citations (5-year frame)	Number of times the focal patent was quoted as relevant state of the art (prior art) during examinations of subsequent patent applications filed within five years after application of the focal patent application	Logarithmic	++: the more often a focal patent is quoted as prior art during examinations of subsequent patent examinations, the more fundamental its technological contribution to the field, the higher its quality	Trajtenberg (1990)
Incremental Forward Citations	Number of times the focal patent was quoted as relevant state of the art (prior art) during examinations of subsequent patent applications filed within the period of five to ten years after publication of the focal patent application	Logarithmic	See above; the effect of the incremental forward cites should, however, play out particularly for fundamental inventions	See above
Accelerated Examination Request (1: yes, 0: no)	Dummy variable taking on the value 1 if a request was filed for an accelerated production of the search report	Binary	+: the higher the applicant's willingness to pay for accelerated protection, the higher the private value of a patent (and, hence, likely the technological quality of a patent) -: the higher the necessity to receive accelerated protection, the more incremental the invention	Reitzig (2004)
PCT I & II (1: yes, 0: no)	Dummy variable taking on the value 1 if a patent was filed via Patent Co-operation Treaty (PCT) in order to seek global protection, and if the period of time between filing date and entry into the regional phase is 20 months or less (PCT I) / exceeds 20 months (PCT II).	Binary	+: the higher the applicant's willingness to invest in global protection for the focal patent (exceeding the EP territory), the higher a patent's commercial value (and, likely, its technological quality) -: the higher the applicant's willingness to pay for the delay of cost-intensive decisions during the application, the higher the applicant's uncertainty about the focal patent's commercial value (and, likely, its technological quality)	Guellec and van Pottelsberghe de la Potterie (2000) For the differences between PCT I and PCT II see Reitzig (2004)

Table 2
Descriptive Statistics

Variable	Mean	Standard Deviation	Minimum	Maximum
Left-hand side variables				
Opposition (1: yes, 0: no)[1)]	0.11		0	1
Rejection of Opposition (1: yes, 0: no) [2)]	0.21		0	1
Amendment after Opposition (1: yes, 0: no) [2)]	0.32		0	1
Revocation of Patent after Opposition (1: yes, 0: no) [2)]	0.34		0	1
Opposition Procedure Closed (1: yes, 0: no) [2)]	0.07		0	1
Opposition Outcome not Definable (1: yes, 0: no) [2)]	0.07		0	1
Exogenous variables (right-hand side)				
Number of Backward Citations to patent Literature (incl. international search)[3)]	2.90	2.44	0	22
Number of Backward Citations to the Non-Patent Literature (incl. international search)[3)]	2.64	2.91	0	41
Number of Designated States (Family Size) [3)]	8.97	3.05	1	13
Number of Applicants[3)]	1.12	0.43	1	7
Number of Inventors[3)]	2.84	1.74	1	19
Number of Forward Citations (5-year frame) [3)]	1.51	2.79	0	37
Number of Forward Citations (10-year frame) [3)]	2.00	3.32	0	44
Accelerated Examination Request (1: yes, 0: no) [3)]	0.02		0	1
PCT I (1: yes, 0: no) [3) 4)]	0.03		0	1
PCT II (1: yes, 0: no) [3) 4)]	0.10		0	1

Legend:
1): Figures calculated for the sample of granted patents comprising N=3,162 patents
2): Figures calculated for the sample of opposed patents comprising N=334 patents.
3): Figures calculated for the entire sample comprising N=5,051 patent (application)s.
4): PCTI and PCT II are distinguished from one another based on the time elapsed between the patent's priority date and the entry into the regional phase. Theoretically, the time lapsed should never exceed 30 months. For a small fraction (8%) of our PCT cases we do observe lags that exceed 30 months, however. We were not able to resolve this puzzle, but we deem it minor for the analysis.

Table 3

Reduced Form Estimates

		Model Grant (N=5027) Stage One: Impact of Patent Quality on Probability of Patent Being Granted relative to Not Granted upon Application **Excluding Information Revealed Largely After Patent Grant**			**Model Oppo** (N=310) Stage Three: Impact of Patent Quality on Probability of Patent Being Upheld relative to being Revoked or Amended upon Challenge **Excluding Information Revealed Largely After Patent Grant**	
K	**Parameter**	**Est. B**	**s.e.**		**Est. B**	**s.e.**
1	Intercept	-0.1407	0.2105		-1.3659	1.2509
2	ln(1+patent references)	0.2480	0.0479	**	0.2226	0.2078
3	ln(1+non-patent references)	0.0346	0.0416		-0.2001	0.1772
4	ln(# designated states)	-0.2276	0.0645	**	0.4397	0.4631
5	Acc. Exam. Request	0.7923	0.2890	**	-0.9651	0.7875
6	Applicant (=1)	0.1216	0.1061		-0.1962	0.4605
7	ln(1+ # inventors)	0.4320	0.0741	**	-0.1757	0.3130
8	PCTI (=1)	0.2765	0.1775		0.8807	0.5726
9	PCTII (=1)	-0.1628	0.1031		-0.6366	0.5606
10a	ln(1+5-year cites)	0.2865	0.0466	**	-0.0861	0.1529
		Log-Likelihood (0): -3484.45; Log-L (model): -3249.09			Log-Likelihood (0): -214.876; Log-L (model): -182.323	

* - significant at the α=.05 level; ** - significant at the α=.01 le

Table 4

Model of CompleteHomogeneity (with Scale/Variance Heterogeneity)
Stages 1 (Patent Grant) and 3 (Opposition Outcome) Pooled

Model 4
Excluding Information Revealed Largely After Patent Grant
(N=5337)

K	Parameter	Est. B	s.e.	t-stat	p-value	
1	Intercept	-0.0718	0.1282	-0.5603	0.5753	
2	ln(1+patent references)	0.1394	0.0289	4.8161	0.0000	**
3	ln(1+non-patent references)	0.0289	0.0251	1.1494	0.2504	
4	ln(# designated states)	-0.1379	0.0396	-3.4822	0.0005	**
5	Acc. Exam. Request	0.5181	0.1693	3.0602	0.0022	**
6	Applicant (=1)	0.0787	0.0643	1.2247	0.2207	
7	ln(1+ # inventors)	0.2634	0.0448	5.8745	0.0000	**
8	PCTI (=1)	0.1169	0.1046	1.1176	0.2637	
9	PCTII (=1)	-0.0813	0.0627	-1.2969	0.1947	
10a	ln(1+5-year cites)	0.1707	0.0275	6.2148	0.0000	**
	Scale Parameter					
	Scale (Stage 1)	1.6117	0.0774	20.8274	0.0000	**
	Scale (Stage 3)	-1.4870	0.2457	-6.0525	0.0000	**

Log-Likelihood (0): -3699.327; Log-L (model): -3436.408

* - significant at the α=.05 level; ** - significant at the α=.01 level

Table 5

Model of Partial Heterogeneity Relating to Post Grant Information (with Scale/Variance Heterogeneity)

Stages 1 (Patent Grant) and 3 (Opposition Outcome) Pooled

Model 5
Including Information Revealed Largely After Patent Grant
(Homogenous intercept)
(N=5337)

K	Parameter	Est. B	s.e.	t-stat	p-value	
	Homogeneous Parameters					
1	Intercept	-0.1401	0.1239	-1.1304	0.2583	
2	ln(1+patent references)	0.1386	0.0276	5.0167	0.0000	**
3	ln(1+non-patent references)	0.0277	0.0244	1.1335	0.2570	
4	ln(# designated states)	-0.1277	0.0378	-3.3761	0.0007	**
5	Acc. Exam. Request	0.4671	0.1703	2.7433	0.0061	**
6	Applicant(=1)	0.0741	0.0626	1.1845	0.2362	
7	ln(1+ # inventors)	0.2899	0.0436	6.6487	0.0000	**
8	PCTI(=1)	0.1624	0.1034	1.5705	0.1163	
9	PCTII(=1)	-0.0981	0.0589	-1.6662	0.0957	
10b	ln(incremental forward cites)	-218.8216^	54.1432	-4.0415	0.0001	**
	Heterogeneous Parameters					
10b	ln(incremental forward cites)	-219.2145^	54.1432	-4.0488	0.0001	**
	Scale Parameters					
	Scale (Stage 1)	1.7017	0.0795	21.3954	0.0000	**
	Scale (Stage 3)	0.0015	0.0004	4.0401	0.0001	**
			Log-Likelihood (0):-3699.327; Log-L (model):-3424.83			

^ - model failed to converge after 1000 iterations; large standard errors/biased estimates indicate overly restrictive model structure.

* - significant at the α=.05 level; ** - significant at the α=.01 level

Table 6

Covariance Heterogeneity Model with Mixture of Heterogeneous and Homogeneous Systematic Assessment Parameters
Stages 1 (Patent Grant) and 3 (Opposition Outcome) Pooled

		Model 6a Excluding Information Revealed Largely After Patent Grant (N=5337)					Model 6b Including Information Revealed Largely After Patent Grant (N=5337)				
K	**Parameter**	**Est. B (lambda)**	**s.e.**	**t-stat**	**p-value**		**Est. B**	**s.e.**	**t-stat**	**p-value**	
	Homogenous parameters										
1	Intercept	-0.5744	0.3296	-1.7426	0.0814		-0.7318	0.4053	-1.8057	0.0710	
2	ln(1+patent references)	0.1808	0.0343	5.2672	0.0000	**	0.2012	0.0397	5.0673	0.0000	**
3	ln(1+non-patent references)	0.0139	0.0298	0.4676	0.6401		0.0273	0.0350	0.7786	0.4362	
4	ln(# designated states)	0.0106	0.1285	0.0822	0.9345		0.0149	0.1569	0.0948	0.9245	
5	Acc. Exam. Request	0.0193	0.2452	0.0787	0.9372		0.0154	0.2977	0.0517	0.9588	
6	Applicant (=1)	0.0739	0.0767	0.9645	0.3348		0.0909	0.0905	1.0041	0.3153	
7	ln(1+ # inventors)	0.1264	0.0912	1.3865	0.1656		0.1591	0.1096	1.4519	0.1465	
8	PCTI(=1)	-0.1354	0.0747	-1.8127	0.0699		-0.1581	0.0854	-1.8498	0.0643	
9	PCTII(=1)	0.0854	0.0417	2.0488	0.0405	*	-0.7318	0.4053	-1.8057	0.0710	
10a	ln(1+5-year cites)	-0.5744	0.3296	-1.7426	0.0814		-	-	-	-	
10b	Ln (incremental forward cites)	-	-	-	-	-	0.2907	0.0785	3.7038	0.0002	**
	Heterogeneous parameters										
1	Intercept	-0.5032	0.3157	-1.5938	0.1110		-0.5612	0.3900	-1.4391	0.1501	
4	ln(# designated states)	0.1810	0.1282	1.4117	0.1580		0.2058	0.1566	1.3137	0.1890	
5	Acc. Exam. Request	-0.5799	0.2452	-2.3655	0.0180	*	-0.6898	0.2977	-2.3167	0.0205	*
7	ln(1+ # inventors)	-0.1972	0.0909	-2.1704	0.0300	*	-0.2750	0.1092	-2.5186	0.0118	*
10a	ln(1+5-year cites)	-0.1309	0.0395	-3.3104	0.0009	**	-	-	-	-	
10b	Ln (incremental forward cites)	-	-	-	-		-0.3003	0.0762	-3.9412	0.0001	**
	Scale Parameters										
	Stage 1	1.3297	0.0639	20.8201	0.0000	**	1.1339	0.0530	21.3894	0.0000	**
	Stage 3	1.7818	0.2456	7.2563	0.0000	**	1.4531	0.2010	7.2299	0.0000	**
		Log-Likelihood (0): -3699.327; Log-L (model): -3426.318					Log-Likelihood (0): -3699.327; Log-L (model): -3403.65				

* - significant at the α=.05 level; ** - significant at the α=.01 level

Appendix A

Table A1
Correlation Matrix of Independent Variables

	I	II	III	IV	V	VI	VII	VIII	IX	X
I: ln(1+patent references)	1.0000									
II: ln(1+non-patent references)	-0.1469	1.0000								
III: ln(# designated states)	-0.0432	0.0745	1.0000							
IV: Acc. Exam. Request	-0.0045	0.0421	0.0287	1.0000						
V: Applicant (=1)	-0.0262	0.0628	0.0160	0.0188	1.0000					
VI: ln(1+ # inventors)	-0.0070	0.0678	0.0030	0.0113	0.1441	1.0000				
VII: PCT I	0.0616	0.1141	0.0203	-0.0136	0.0269	0.0105	1.0000			
VIII: PCT I	0.0892	0.1563	0.0215	0.0062	0.0119	-0.0649	-0.0615	1.0000		
IX: ln(1+5-year cites)	0	0.2463	-0.0000	0.0546	-0.0149	0.0074	0.1194	-0.1165	-0.2556	1.0000
X: Ln (incremental forward cites)	0.2167	-0.0411	0.0203	-0.0097	-0.0068	-0.0020	-0.0886	-0.1767	0.3730	1.0000

Table A2
Prediction of patent grant based on models Grant
(N=5,027 after outlier correction)

	No Grant (Predicted)	Grant (Predicted)
No Grant (Real)	263	1,626
Grant (Real)	203	2,959

Appendix B: A More Elaborate Description of Variance Decomposition Discrete Choice Models

Using the basic axioms of Random Utility Theory (RUT), unobservable (latent) technological quality (=techno-economic quality), LQ_i, of patent 'i' can be expressed as an additive function of its systematic/explainable technological quality, Q_i, and some random/unexplainable component, ε_i. That is,

$$LQ_i = Q_i + \varepsilon_i \quad (1)$$

Systematic technological quality (Q_i) is assumed to be a generalized regression function of various observable and measurable factors. In turn, these factors ultimately determine the overall technological quality of a patent as judged by a patent office and hence its likelihood of being granted upon application or upheld upon challenge. We assume this function to be linear in the parameters (Ben-Akiva and Lerman 1985). We define a matrix $\mathbf{X}_i$ which describes the measurable technological quality of patent 'i' on various attributes (see Table 1, exogenous variables). We define a set of parameters, $\boldsymbol{\beta}$ which capture the effect that these factors have on changes in mean (systematic) technological quality. In general, the impact that each dimension of a patent application has on its mean technological quality is:

$$Q_{is} = \mathbf{X}_i \boldsymbol{\beta}_s \quad (2)$$

We use the subscript 's' to suggest that the perceived quality of a patent at different stages of a patent process (initial application; patent opposition) may be different. In particular, while the (observable) patent characteristics may be constant, it is possible that the average correlation of these characteristics with its perceived quality may differ from stage to stage; hence, requiring a separate set of parameters, $\boldsymbol{\beta}$, for each stage 's'.

Among the basic empirical models capable of estimating different sets of parameters for similar outcomes using discrete choice data are so-called multinomial models. Essentially, by comparing sets of different parameters, $\boldsymbol{\beta}$, for each stage 's' in our data structure allows us to assess the

consistency of the decision making across the different stages. This is why we use discrete choice models, however, there are various challenges in applying discrete choice modeling to our data.

McFadden (1974) introduces several axioms to construct the (basic) multinomial logit model, including Independence-from-Irrelevant Alternatives (IIA), positivity, and irrelevance of alternative set effect. This implies that the random elements, ε, are iid. By further assuming that this distribution is Gumbel (extreme value type I), the closed-form MNL can be constructed (Ben-Akiva & Lerman 1985). In the multinomial logit model, one assumes that the error component, ε, is distributed iid Gumbel, with a zero location parameter (without loss of generality) and scale parameter, λ. Applying a multinomial logic to our data, from McFadden (1974), the probability that a patent application 'i' ends up in one of 'J' scenarios, at observation 't', at stage 's' of the patent application process can then be expressed as:

$$P_{its} = \frac{\exp(\lambda_s Q_{its})}{\sum_{j=1}^{J} \exp(\lambda_s Q_{jts})} \quad (3)$$

In the first stage (s=1), P_{it} is the probability that patent 'i' will be granted. The technological quality required for a patent not to be granted must be set to some threshold value (e.g. zero) for identification purposes, consistent with a binary logistic regression expression. In the opposition ("challenge") stage, we bundle the three outcomes (rejection of opposition/amendment of patent/revocation of patent) into two, and P_{it} is the probability that patent 'i' is maintained as granted (rejection of opposition) or not (patent amended or revoked) (note: by bundling the outcomes in this fashion, the thresholds for technological quality in Stage 1 and Stage 3 are set equal). The technological quality threshold of a patent being revoked or amended upon such challenge is set to zero, again resulting in a binary logistic expression. In turn, concerns about IIA violations are not applicable, contrary to what one might think at first sight when looking at the "nested" structure of our data.

In this model, it is not well known that the estimates of vector β, of length 'k', describing the impact of various factors on mean systematic quality, are confounded with scale (Louviere, 2001). In any single data set, the scale parameter of the random component, λ, a scalar, is not identifiable, so the usual procedure is to arbitrarily set its value to 1. By ignoring this parameter, however, one could

make erroneous conclusions about the true assessments of technological quality by the patent office in the different stages (Stage 1: grant yes/now; Stage 3: patent revoked/amended/upheld as granted). Specifically, conclusions about differences between decisions based on estimates of β could be explained by differences in the true underlying assessment structure with respect to technological quality, differences in underlying variability or both (Louviere, 2001).

When one estimates a single discrete choice model with a latent dependent variable (including probit, mixed logit), true estimates of β are confounded with the scale parameter. In turn, the estimates are ($\lambda\beta$), where λ is the scale parameter associated with that particular set of data. The scale is inversely related to the variance of the random component, σ_ε^2 by the relation:

$$\lambda = \sqrt{\frac{\pi^2}{6\sigma_\varepsilon^2}} \qquad (4)$$

In turn, when we compare parameter estimates related to systematic components of technological quality assessments, we actually compare a confounded set of parameters. For instance, estimates describing the impact of patent characteristics on the likelihood of a patent being granted (Stage 1) may be ($\lambda_1\beta_1$). Estimates describing the impact of patent characteristics on the likelihood of a patent being upheld upon opposition (Stage 3) may be denoted ($\lambda_3\beta_3$). Although we often arbitrarily set the value of λ_1 and λ_2 to unity (as most statistical packages do), we cannot be sure that in comparing estimates from two models, say ($\lambda_1\beta_1$) to ($\lambda_3\beta_3$), that differences are differences in true underlying technological quality assessments (i.e., heterogeneous β), differences in the variance of the random components (i.e., heterogeneous λ), or simultaneously differences in both sets of parameters (this problem was first addressed in the quantitative marketing literature by Ben-Akivaet.al. 1994, Hensher, Louviere, & Swait, 1999, and Louviere, Fox, & Moore, 1993). [34]

[34] In other areas, the issue related to comparing confounded estimates has been noted and used to identify that erroneous conclusions may have often been made in ignoring this statistical truth. For instance, in marketing science several authors have demonstrated empirically that often differences that appear to be occurring in terms of consumer preference observed in real markets relative to preferences obtained in hypothetical settings (e.g., choice experiment) can be dismissed once the differences in variability in consumers choices across the two settings are accounted for (Ben-Akiva et al. 1994; Hensher, Louviere, and Swait 1999; Louviere, Fox, and Moore 1993). For example, the way in which consumers trade-off price (e.g., prefer products with lower prices) and aspects of quality (e.g., prefer higher quality products) is often the same whether these evaluations are made in relation to real products (i.e., revealed preferences) or in relation to hypothetical products (i.e., stated preferences). In turn, once accounting for

Essentially, the underlying empirical possibilities that require testing are two fold. Firstly, we wish to ascertain whether technological quality assessments by a patent office are the same (i.e., $\beta_1 = \beta_2$) across granting and challenge stages. Second we wish to do this but account for whether the scale parameters (corresponding to variability) are also the same (i.e., $\lambda_1 = \lambda_2$). Essentially, we wish to test whether assessments of a patents technological merit relate to several observable characteristics of this patent and consider if these systematic assessments are homogeneous or heterogeneous across two data sets (or populations); we simultaneously test whether the randomness with which choices are observed are heterogeneous or homogeneous across two data sets.

In order to address this issue, Swait & Louviere (1993) propose a nested hypothesis testing procedure.

First, the authors estimate a model in which complete heterogeneity is imposed on both the scale and assessment components relating to technological quality. That is, essentially, a sample-specific set of parameters is estimated for each stage (Stage 1: all patent applications; Stage 3: opposed patents only). In order to do this, however, the scale parameter cannot be identified and set arbitrarily to one in any one data set. The model log-likelihoods, however, provide a base measure for which subsequent models imposing various aspects of homogeneity can then be compared.

Second, Swait and Louviere propose a model of complete technological quality assessment and variance homogeneity, in which the data from the two samples are pooled (Stage 1: all patent applications; Stage 3: opposed patents only). This model is tested against the base model of complete heterogeneity using a likelihood ratio test.

Third, the authors introduce a model of complete quality assessment homogeneity while relaxing the assumption of variance homogeneity. To do this, they manually multiply the independent measurable components of one data set by a scale ratio and assume the alternative data set has a scale ratio of one. The concavity of the likelihood function with respect to the varying scale ratio allows a maximizing scale ratio to be identified under a hypothesis of assessment homogeneity regarding

---

differences in the variability inherent in these evaluations, it can often be concluded that trade-offs are identical rather than an initial hypothesis that preferences differ.

technological quality. Comparing this model's likelihood to the base model of complete homogeneity allows this hypothesis to be formally tested.

Swait & Louviere's (1993) model only assesses the variance differences (using a manual grid search), given that homogeneity in the systematic component has been established. If, however, the model that allows for variance heterogeneity is significantly different from the model of complete homogeneity one question remains. Namely, is this what drives the heterogeneity of the variance? We use our own purpose-written software which allows this to be estimated using a full-information maximum likelihood (FIML) approach in which a Newton-Raphson algorithm is used, given a closed form solution. In more detail, we propose to model and test the variation from an assumption of homogeneity by estimating two parameters for each of the potentially heterogeneous variables, such that one parameter (homogenous term) describes the average impact across assessment Stages 1 (Grant) and 3 (opposition outcome) and another parameter describes the deviation from this average impact for one assessment stage relative to the other. We use τ to denote the sets of factors suspected of being considered by patent offices in a heterogeneous fashion across the two assessment stages. We let the impact of these factors be determined by:

$$\beta_\tau = \beta_\tau^* + \beta_{\tau s} Z_s \qquad (5)$$

where Z_s = -1 if initial stage and +1 if challenge stage. β_τ^* is the average impact of factor τ, and $\beta_{\tau s}$ provides a test of the degree to which such homogeneity across stages is being violated. That is, the impact of factor τ for the initial stage is given by:

$$\beta_{\tau\, initial} = \beta_\tau^* - \beta_{\tau s} \qquad (6)$$

and the impact of factor τ for the challenge stage is given by:

$$\beta_{\tau\, challenge} = \beta_\tau^* + \beta_{\tau s} \qquad (7)$$

In turn, the t-statistic associated with the mean estimate of $\beta_{\tau s}$ provides a formal test to assess whether an assumption of preference homogeneity is significantly violated (i.e., Ho: $\beta_{\tau s} = 0$). Since each model is also nested within the previous models of homogeneity, appropriate likelihood ratio tests are applicable to further confirm the resulting model.