MERITOCRACY VOTING:
MEASURING THE UNMEASURABLE

by

Peter C. B. Phillips

COWLES FOUNDATION PAPER NO. 1504

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2016

http://cowles.yale.edu/

Econometric Reviews, 35(1):2–40, 2016

ISSN: 0747-4938 print/1532-4168 online
DOI: 10.1080/07474938.2014.956633

Meritocracy Voting: Measuring the Unmeasurable

Peter C. B. Phillips

Yale University, New Haven, Connecticut, USA; University of Auckland, Auckland, New Zealand; University of Southampton, UK; and Singapore Management University, Singapore

Learned societies commonly carry out selection processes to add new fellows to an existing fellowship. Criteria vary across societies but are typically based on subjective judgments concerning the merit of individuals who are nominated for fellowships. These subjective assessments may be made by existing fellows as they vote in elections to determine the new fellows or they may be decided by a selection committee of fellows and officers of the society who determine merit after reviewing nominations and written assessments. Human judgment inevitably plays a central role in these determinations and, notwithstanding its limitations, is usually regarded as being a necessary ingredient in making an overall assessment of qualifications for fellowship. The present article suggests a mechanism by which these merit assessments may be complemented with a quantitative rule that incorporates both subjective and objective elements. The goal of "measuring merit" may be elusive, but quantitative assessment rules can help to widen the effective electorate (for instance, by including the decisions of editors, the judgments of independent referees, and received opinion about research) and mitigate distortions that can arise from cluster effects, invisible college coalition voting, and inner sanctum bias. The rule considered here is designed to assist the selection process by explicitly taking into account subjective assessments of individual candidates for election as well as direct quantitative measures of quality obtained from bibliometric data. Audit methods are suggested to mitigate possible gaming effects by electors in the peer review process. The methodology has application to a wide arena of quality assessment and professional ranking exercises. Some specific issues of implementation are discussed in the context of the Econometric Society fellowship elections.

Keywords Auditing peer review; Bibliometric data; Election; Fellowship; Measurement; Meritocracy; Peer review; Quantification; Subjective assessment; Voting.

JEL Classification A14; Z13; C18.

"Man must not be afraid of what seems impossible to do. History has shown that human beings possess a wonderful gift of being able to obey the saying of

Address correspondence to Peter C. B. Phillips, Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520-8281, USA. E-mail: peter.phillips@yale.edu
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lecr.

Aristotle: 'Measure the Unmeasurable.' " Ragnar Frisch (Examination report as a student at the University of Oslo, cited in Andvig and Thonstad, 1998, and later by Louça, 2007.)[1]

1. INTRODUCTION

Hierarchical elements and status inequalities are pervasive in modern industrialized society. Social stratifications arise from multiple sources such as socio-economic conditions, occupation or profession, earnings, and education. Affiliation with the military or religious orders affects community status just as industrial power, media exposure, and political influence enhance visibility in society. By contrast, anthropologists argue that some hunter-gathering societies are (or were) relatively free from social stratification. Those societies typically comprised small acephalous (or headless) tribal foraging groups where tasks were more uniformly distributed across a group and decision making was largely by consensus and there were fewer societal distinctions (Gowdy, 2006).

When stratifications do exist in society, distinctions are usually clear enough to identify groupings of individuals according to certain characteristics such as income and influence. Quantitative measurement can be straightforward in some categorizations, but qualitative assessment is often needed in others. Categorical information helps in distinguishing groups like Fortune 500 companies and celebrity billionaires, and in providing classifications such as senior or middle management in industry; quantitative data provide fine grain information on a myriad of detail concerning characteristics such as income, wealth, age, size of family, years of education, and so on.

Learned societies, which are the focus of the present work, also operate stratified social structures. These societal structures form a meritocracy in which some members occupy elevated positions relative to others, at least for a time. Virtually all learned societies have presidents as leaders, a governing body or council that determines policy, and an executive committee or officer(s) as an administrative arm—all with fixed terms. Many societies award fellowships—usually for life—to members whose credentials distinguish

[1]The original source was Frisch's examination article in public finance in 1919, a document that has survived but has never been published. Olav Bjerkholt has provided the following literal translation of the original:

> "Man must not be deterred by the apparently impossible. History has shown that the human beings have had a wonderful ability for obeying the maxim of Aristotle: 'Make the unmeasurable measurable.' "

These words appear to be the first words written by Frisch to be published. Remarkably for an examination article they formulate, as Andvig and Thonstad (1998, p. 6) put it, "his own overriding future research policy" to tackle the apparent impossible. Much of the early empirical econometric work on measuring demand elasticities (e.g., Schultz, 1924) may well have appeared in the same light at that time.

them within the society. Some also offer distinguished fellowships which honor lifetime contributions to a discipline. Such fellowships offer status and lead to a stratified structure of membership within a society that becomes a distinguishing characteristic of its meritocracy. Fellowship in a leading international society is generally considered to be a singular honor. As a public endorsement of merit and accomplishment, it can have a lasting effect on a career and remuneration. Accordingly, it is highly prized.

The subject of the present article is the selection process by which such fellowships are determined. Assessment of merit necessarily involves human judgment about the contributions of individual candidates. But information about and opinions of those contributions may differ considerably in a voting population. Part of our goal is to confront the analytic challenge of combining relevant information and opinions in a way that assists the overall assessment of candidates in a meritocracy vote. Our approach is to construct a methodology for combining "objective" and "subjective" information for use in such voting and to broaden the availability of that information during the voting process. Modern webserver facilities afford community access to massive datasets that can be tailored to deliver specific information requirements to assist voters in decision making. We outline a methodological framework to fortify peer review with such information. While some examples are given, this is not an article on specific bibliometric or citation measures. There is now a vast and growing literature, with many experts, on that subject. This literature is important to the mechanics of quality measurement and to alert us of the strengths and limitations of the multitudinous measures that are now available. The present work has a different orientation. Its focus is directed toward honor society voting and the analytic mechanisms for building more information into that process rather than specific details of the information to be deployed and how that might evolve as more data become available.

A secondary goal of the article is to open up public discussion amongst economists and econometricians of the issues involved and how these may affect our academic societies amidst an explosion of ranking data on individuals and institutions. A consensus in such discussion may seem out of reach. But the econometrics profession is well positioned to promote the advancement of evidence-enhanced voting procedures and suggest mechanisms for incorporating such evidence. The ensuing discussion of this article and later research by other econometricians may usefully widen the focus to suggest details of the measurement mechanics, including the bibliometric and citation data that may be mobilized, with attendant caveats, in the voting process.

Debate on the qualifications for fellowship is as ancient as learned societies themselves—witness the famous Hobbes–Wallis controversy involving the Royal Society in the 17th century. In an archival study on the foundation of the Econometric Society, Bjerkholt (1998) recently provided extensive evidence of diverging views among the founders of that Society in the early 1930s about electoral procedures and about individual candidates for fellows. In a further study, Louça and Terlica (2011) report continuing divisive debates among the broader fellowship in the 1950s about selection criteria for fellowship.[2] The issues that manifested in these Econometric Society (ES) debates relate largely to field qualifications. Concerns over field remain unresolved in the 21st century. They have focused on growing field disparities in Econometric Society fellowship elections and societal appointments as well as the role within the ES of econometrics itself.[3]

Distortions in voting may arise for many reasons. For instance, intellectual founders and leaders may veto certain candidates[4]; and coalitions of voters can form among visible (i.e., physically extant) and invisible (e.g., by subfield or intellectual descent) colleges of electors to secure election for preferred candidates. How, in such a system, can the merit that underlies a meritocracy be fairly determined, or even defined? What elements—quantitative and qualitative—might enter into the selection process to substantiate

[2]The 1950s debate was prompted by correspondence of Oscar Morgenstern circulated in 1953 to all fellows of the Econometric Society stating that

> in my view the Fellows ought to be persons who have done some econometric work in the strictest sense. That is to say, they must have been in one way or another in actual contact with data they have explored and exploited, for which purpose they may have even developed new methods.

This viewpoint was strongly supported by some fellows (among them Robert Geary, Charles Roos, and P. C. Mahalanobis) and opposed by others (including Tjalling Koopmans and Jacob Marschak). In the end, no changes to criteria or procedures for fellowships were made.

[3]In a letter to the President of the Econometric Society on June 26, 2010, the author and David Hendry raised concerns about the role of econometrics within the ES, pointing to

> a mounting concern that the Econometric Society has become progressively less representative of econometricians within the society with consequential impacts, particularly on the careers of younger econometricians. To many there is an emergent crisis in econometrics because of the lack of acknowledgement and representation and the growing difficulties econometricians have in publishing in *Econometrica* and other general interest journals in economics. The movement away from econometrics is manifested each year in the election of officers, council and fellows, the appointment of Editors of *Econometrica*, and recently by the formation and nature of the new journals. While the concern over under-representation has occasionally been raised in Fellows Meetings at various Econometric Society conferences since the 1980s, the situation seems to many to have grown considerably worse over the past decade. Many people are now puzzled about the role of the Econometric Society in terms of what it does for econometrics and the increasing lack of congruence between its name and its focus. By contrast and in response to the direct needs of the econometrics community, a large number of highly successful regional and thematic meetings have been organized that are outside the aegis of the Econometric Society and continue to grow and prosper without any connection to or support from the Econometric Society.

[4]From archival research on correspondence among the Council of the Econometric Society in the early 1930s, Louça (2007) reports that one candidate for a fellowship was opposed on the grounds that "he would not know a partial derivative" (p. 31), an injustice as it turned out. Bjerkholt (1998, p. 53 and footnote 32) provides original source material and further details on this particular incident. Another candidate was repeatedly opposed as President of the Society as "not recommendable" on the grounds that he "uses many words to express his meanings" (Louça, 2007, p. 35).

election? If democratic voting is involved in the selection, how might the human electorate (of voters) and individual motives be complemented with a material electorate (of data) so as to promote informed and fair election that mitigates potential distortions? How, in short, may weaknesses in the democratic voting system be attenuated in societal decisions on merit?

In empirical research, Hamermesh and Schmidt (2003) analyzed data from fellowship elections in the Econometric Society over the period 1990–2000 to assess whether these elections were "fair" in the sense that the votes cast accorded with candidate qualifications. Objective measures of quality were based on (i) the average number of citations to the candidate's work over the two preceding years, (ii) a count of the candidate's publications in *Econometrica* (the Econometric Society's journal), and (iii) an indicator of whether the candidate had ever been an Associate Editor or Coeditor of *Econometrica*. Controlling for this measure of quality, logit and probit regressions were used to assess the empirical significance of various other determinants of the election outcomes. The results revealed that successful election depended on many characteristics other than quality, including current affiliation, field, and geographical location.

All voting systems are subject to potential gaming decisions by electors. For instance, in meritocracy voting where there are thresholds and quotas for election, individual elector decisions to support, abstain, or rank candidates can end up having a major impact on outcomes. Coalitions among electors can accentuate this impact, as intimated above. They can arise by explicit or implicit agreement, possibly from a dimension of commonality such as institutional or field affiliation. Top tier institution bias is an example. Leading institutions often have a large concentration of electors because of the number of fellows already working in the institution. New candidates for election from within the institution then have an advantage over other candidates due to (i) extended common knowledge within the institution of the candidate amongst existing fellows, which leads to enhanced cross field voting in support of such candidates, and (ii) pressure to sustain or raise perceived institutional status by electing new fellows from within the institution. The latter pressure sometimes takes the form of explicit exhortations by senior management, deans, and department chairs to elect new fellows from the professoriate in order to help raise the institution's profile. Similar pressures can operate within countries, regions, subfields, or invisible colleges of electors with common academic pedigrees.

Societies where fellowship decisions are made in committee (such as the Institute of Mathematical Statistics and the American Statistical Association) rely on peer evaluation within these committees in reviewing candidate nomination materials, letters of reference, publication records, research papers, and other evidence such as citation data and records of mentorship. Subjective assessments of candidates may then be presented with this supporting evidence, and individual cases can be discussed and decided in a process that mirrors committee-based promotion decisions and appointment processes in universities. At a narrower level, this process is analogous to professional journal review where referee evaluations are solicited in conjunction with associate editor or co-editor recommendations. Even when such committees conduct formal votes on candidates, there is less opportunity to game the final decision in this system because of transparency within small committees. On the other hand, committee decision making is highly subjective, and the information set that affects decisions is limited to the material presented and by the knowledge base of the committee members and any research that they may do on candidates.

In both systems, greater use of quantitative data can enhance informed decision making. There is also considerable scope for using crowd wisdom within the profession to raise awareness of the strengths of less well-known candidates, those working outside the major centres of learning, and those working in less populated or emerging subfields.

Finding a mechanism for promoting fairness across fields, institutions, and regions, collecting and distributing the relevant information that can assist in this process, and respecting subjective assessments of credentials across a population of electors are serious challenges for any society. Societies in quantitative disciplines like economics may well be expected to rise to this challenge, as Frisch enjoined in the header to this article, and show leadership in creating and testing such selection mechanisms.[5]

This article seeks to offer some material assistance toward that goal and to open the issues up for professional discussion so that the best ideas may be taken forward. Our work here provides a quantitative rule that combines human judgment and quantitative data on credentials in a mechanism that brings this disparate information into the election or selection process without removing the effect of individual votes on the outcome of a candidate's election. The goal, in short, is to assist the process of voting on merit by *measuring merit* — measuring the unmeasurable—by widening the effective electorate that enters the decision process with a broad additional class of objective and subjective elements. These elements involve a comprehensive (i.e., electorate wide) peer evaluation component that is combined with bibliometric measures to determine an explicit merit threshold (a vote percentage) that is needed for election. Peer review and individual votes continue to play a key role, but they are complemented with material evidence on accomplishment.

The statistical use of bibliometric data in combination with comprehensive peer assessment has many potential applications that extend beyond the immediate arena of fellowship elections. Research assessment exercises that are now undertaken in some countries (such as the U.K., Australia, and New Zealand) are one example where such

[5]Some journals in economics already use automated measures in determining fellowships (*Journal of Econometrics*), distinguished authorships (*Journal of Applied Econometrics*), and annual prizes (*Econometric Theory*). The measures employed in these awards rely on bibliometric counts and are not complemented with peer review data.

data may be used.[6] Journal rankings and impact factors of research are another.[7] Senior management teams of universities and journal publishers now make substantial use of such credentials in promoting their institutions and publications. Researchers who are accustomed to peer review processes in journal and promotion decisions often find themselves uncomfortable with the mechanical approaches that are typically adopted in producing these rankings, especially when they are obtained by automated harvesting of bibliometric or citation data and search engine methods that are themselves subject to measurement error and outlier effects. The challenge we face in such assessment exercises is to utilize the vast and growing quantity of such data in a manner that complements established peer evaluation processes which most professionals view as a necessary component in quality assessment. The methodology explored in the present article provides a mechanism to address that challenge and strengthen the data-based foundation of the quality assessment process.

2. MERIT THRESHOLD AND CREDENTIALS

In societies where fellowship elections are held, candidates need to achieve a certain threshold percentage (τ) of positive votes from the electorate of voters to be successful. This voting electorate might be the collection of all existing fellows in the society, a fellowship selection committee, the governing body or council, or even the entire society membership. Examples from leading learned societies in economics, statistics, and national academies are collected and discussed in the Appendix. In what follows we will concentrate on developing a mechanism that is suited to a wide-body voting electorate such as all existing fellows, as in the present system of the Econometric Society, or full society membership voting. Some fairly obvious modifications to suit other voting electorates can be made in the system that is described below.

The threshold percentage τ may be arbitrary, such as some number in a certain interval like $\tau \in (0.25, 0.75)$, and it might be set by the governing body of the society or the selection committee for voting decisions in a committtee on new fellows. Underlying τ—either explicit or implicit—is a social welfare function view concerning the size of the fellowship as a proportion of societal membership. That view may be determined in the articles of association or the constitution of the society or it may be a matter on which the council or membership votes from time to time, taking into account issues such as

[6]Even when bibliometric data are not explicitly used in the preparation of research assessment exercises or other official rankings, that data may already appear indirectly through individual use in providing "peer esteem" evaluations that many ranking reports request. The use of automated software that universities now provide for harvesting publication information also implies that bibliometric data are being routinely utilized in report preparation.

[7]See Chang et al. (2012) for a recent overview of various bibliometric research assessment measures and an empirical application of the methodology to journals in econometrics and statistics using journal citation reports and the Thomson Reuters ISI Web of Science database.

generational balance and the future composition and size of the society and its fellowship. In such cases, it is known that voting over size gives rise over time to decisions that are time inconsistent (Roberts, 1999). In the Econometric Society, the voting electorate is the body of existing fellows and the threshold is set to a common value with $\tau = 0.30$. In a society with over 450 active fellows, to be successful in an election where all fellows vote a candidate must secure votes from at least 135 fellows. Thresholds may be accompanied or replaced by a quota system for new fellows, as in the case of the Royal Society where the quota is currently 44 new fellows each year, or the National Academy of Sciences where the quota is 84 new members annually.

Thresholds are often decisive in elections. If many strong candidates fall short of attaining the required percentage of favorable votes, a societal governing body may adjust the threshold downwards to increase the number of successful candidates in subsequent elections—possibly to improve perceived social welfare or generational balance. If the threshold is considered too lenient, then it may correspondingly be increased. In this sense, the threshold is endogenous. Its value may be reactive both to past election results and to governing body opinion regarding exclusivity. In effect, the number τ is a voting merit threshold for fellowship which relies directly on inner sanctum views of exclusivity and indirectly on views of past election results. Typically, τ is a common value that applies across all candidates.

The mechanism suggested in the present article seeks to bring further information to bear on this critical merit threshold, to provide a flexible data-based method for the determination of τ, and to make τ individual specific. The mechanism can be used to complement existing systems of election by simply importing information into τ, thereby making the endogeneity of τ explicit and specific to an individual candidate, without removing the power of the human electorate of voters to elect.

The credentials that define merit are subjective and inevitably rely on personal judgment. But they also rely on knowledge (if only by hearsay or on information transmitted in nominating statements and referee reports on candidates) of material accomplishments and personal assessments of the importance and relevance of those contributions. We therefore propose that the merit threshold be determined to explicitly incorporate such information—both objective and judgmental—and to do so in a way that reflects a wide body of base knowledge in the profession arising from published research and its adjudged merit. Importing quantitative and qualitative information in this way widens the effective electorate beyond the immediate voters: for example, published research reflects decisions taken by editors and the judgments of independent referees on the worth of a candidate's research; and citations or online downloads reflect received interest about the research amongst a broad readership of fellow researchers. The goal, in effect, is a mechanism that assists in measuring the “unmeasurable” element of merit in a meritocracy.[8]

3. A FELLOWSHIP ELECTION FORMULA

In what follows, we lay out an evidence and peer review based approach to determine τ. As indicated, we seek to make τ individual specific so that its value may reflect the merits of an individual candidate as measured by the information set that is used in its determination. The distribution of τ across the candidates depends on the distribution of the inputs of objective and subjective information about those candidates for election. The resulting distribution differentiates candidates according to their revealed merit, but it leaves to voters the ultimate task of determining election.

The specific formula given below is parameterized, and the particular choice of parameters will influence outcomes. The formula may be trialed on past election data to find parametric values that correspond closely to actual election outcomes and those that produce alternative results with greater or lesser numbers of successful candidates. In the absence of available empirical data,[9] we will instead report some exact distributional results that show the response distribution of τ to its inputs for certain explicit distributions of objective and subjective evidence These distributions reveal the flexibility of the approach and the way different types and levels of credential information contribute to outcomes.

Quantifying Individual Merit Thresholds

The starting point is to make the merit threshold τ individual specific. In particular, for each nominee a personal threshold of voting support—the merit threshold for that individual—is determined for this person’s election. The merit threshold depends on accomplishment and is measured by an accomplishment factor $X \in [0, 1]$. The factor X is the sum of two components $X = X_a + X_b$, where X_a reflects objective information and X_b embodies judgmental views of the accomplishment. What follows is one possible

[8]To be measured a quantity must first be defined, a task of longstanding relevance in economics whether the quantity in question is the price level, utility, happiness, or merit. In all such cases the inherent multidimensional nature of the quantity (and in the present case the subjective elements involved in defining merit) must be addressed even when a single index proxy variable is used in the accounting process. More subtle is the fact that, however difficult the accounting definition may be, it is still far easier than defining an appropriate probability space for “measuring” the said quantity and using it in statistical work. Almost always in economics and other social sciences, the probability space is too limited for there to exist any “true” representation of the measure (or true data generating process (dgp)), a problem that is still largely untouched in econometrics (see Phillips, 2005, for further discussion).

[9]A request for data on the annual Econometric Society fellowship election outcomes over 1990–2011, including the votes received by each of the nominated candidates, was denied. These data would enable empirical research on ES data of voting formulae such as those considered in this article, the effects of various parameterizations, and some counterfactual analysis.

formula for the determination of X and the manner in which X determines τ. The resulting mechanism inevitably involves some arbitrary elements of construction and specific parameter settings need to be employed to make the formula operational. Later in the article we provide some computations to illustrate the use of this formula and detail its possible implementation. In practice, parameter settings which govern the formula can be set by a society's governing body and modified as may be needed to take account of the evolution of a discipline over time and the views of the society regarding qualifications for fellowship election. The system below is designed in a way that facilitates automated online implementation.

Objective information: determining X_a

The component X_a depends on quantitative information about research accomplishment and material contribution to the discipline. We provide here only a skeleton of ideas to illustrate how X_a might be constructed. One example that we use below is that the governing body may designate certain core journals from which publication data is collected. These might comprise major general interest journals and leading field journals. Sole authored and co-authored publications might be distinguished and weighted in a ratio such as $\rho : 1$ for some relativity parameter ρ. In this case, we may define $Y = \rho n_1 + n_2$ as the core journal publication component where n_1 is the number of sole authored publications and n_2 the number of co-authored publications. In the illustration that follows, we set $\rho = 2$ for simplicity, and extensions to the general case are straightforward. Publication numbers beyond some limit (M) may be ignored in order to delimit quantity effects. Then, defining $N = \min(Y, M)$, the "objective" data component X_a may be constructed as $X_a = \frac{1}{2} \times \frac{N}{M} \in \left[0, \frac{1}{2}\right]$. In a similar manner, X_a can be modified to take into account citations and other data-based measures of research performance and impact. Since such extensions are fairly obvious and may be individually weighted as components of X_a, they will not be explored here. The idea is clear enough, even though the details of determining which indices to use and how they might be constructed will inevitably be more complex given the vast literature that now exists on the subject of citation indices and rankings.

Importantly, we confine the support of the objective component X_a to a fixed subinterval U_a of $[0, 1]$, leaving a residual subset for subjective assessment. With the specific rule $X_a = \frac{1}{2} \times \frac{N}{M}$, the support $U_a = \left[0, \frac{1}{2}\right]$ and X_a carries an implicit weight of $\frac{1}{2}$ in the overall measure X. This weighting system can be altered to reflect a societal view concerning the importance of quantitative information relative to subjective assessment, as discussed further below.

Peer evaluation: determining X_b

The component X_b measures the electorate's collective peer evaluation of a candidate's qualifications for election. There are various ways in which X_b may be determined. We start with the following basic approach, which presumes honesty in peer evaluations with no gaming of the system by electors. The basic system is then expanded to control for gaming effects.

Each member (j) of the voting electorate reports an honest subjective assessment factor $f_j \in [0, 1]$ of the candidate (with higher values of f denoting higher subjective assessment on the $[0, 1]$ scale). To simplify notation at this point, we will use an index (i) for the candidate. The assessments are averaged to produce a subjective accomplishment factor $f = \frac{1}{\#(S_{\text{all}})} \sum_{j \in S_{\text{all}}} f_j$ where S_{all} is the set of all voters (e.g., existing fellows) in the electorate. Precise rules may be given for determining f_j in the case of abstentions, no returns, or invalid returns. For example, if $\varphi_j \in [0, 1]$ is the subjective assessment of the candidate by elector j, we may determine f_j as follows:

$$\begin{aligned} f_j = {} & \varphi_j \times 1\left\{j \text{ returns a subjective assessment factor } \varphi_j \in (0,1)\right\} \\ & + \varphi \times 1\left\{j \text{ abstains, does not vote, or returns a } \varphi_j \notin (0,1)\right\}, \end{aligned} \tag{1}$$

where

$$\varphi = \varphi_{\text{valid}} = \frac{1}{\#\left(S_{\text{valid}}\right)} \sum_{k \in S_{\text{valid}}} \varphi_k,$$

and S_{valid} is the set of electors who returned a valid assessment factor $\varphi \in (0, 1)$. According to this rule, abstentions, nonvoters and extreme assessments $\varphi_j \notin (0, 1)$ are eliminated and replaced by the average peer assessment (φ) over all those electors returning a valid assessment. An alternative rule which assigns greater weight to the electors who nominated the candidate for election would determine φ as

$$\varphi = \varphi_{\text{nom}} = \frac{1}{\#\left(S_{\text{nom}}\right)} \sum_{k \in S_{\text{nom}}} \varphi_k, \tag{2}$$

where S_{nom} is the set of electors who nominated the candidate and returned a valid assessment factor $\varphi \in (0, 1)$ for this candidate.

In both these rules, extreme $0, 1$ assessments are taken to be invalid. This device is merely suggestive at this point, and a more realistic version for implementation is given below. The idea is to force electors to think more carefully about fractional assessments to mitigate the effects of extreme positions. Just as the upper limit M controls tail event effects in X_a by truncation (winsorizing the data), extreme subjective assessments may be controlled in X_b by adjusting the support of φ. The extended system given below shows how subjective extremes may be controlled by the threat of auditing the peer review.

Differential information about candidates is an inevitable factor influencing peer assessment and votes. Field differentials may be so great that electors have difficulty appreciating or even reading the work of candidates outside their own fields. In such cases, honest electors may not be comfortable returning a personal peer review of the candidate, but may be ready to delegate their assessment to others, such as (i) the nominating group, (ii) the fellows selection committee (if one exists), or (iii) other better informed fellows. Assignation to these alternatives may be arranged by an elector being offered distinct discrete choices that signal these reassigned assessments, which might be triggered online automatically by the elector choosing certain discrete integers such as (i) $\varphi_j = 2$, (ii) $\varphi_j = 3$, and (iii) $\varphi_j = 1$ or 0, as in rule (1).

Discrete choices for reassigning assessments might be accompanied by more explicit self-selection by electors about their own expertise in a field. For instance, electors might be asked to rank themselves into three general categories as (i) expert (R_1), (ii) informed but not expert (R_2), or (iii) an outsider (R_3) to the research field of each candidate. These expertise self-selection outcomes, denoted $\{R_{jia}, a = 1, 2, 3\}$ for candidate i by elector j, might then be used to weight electors subjective assessments according to some agreed system (e.g., $w_{ia} = 1, \frac{1}{2}, \frac{1}{4}$ for expert, informed, and outsider, respectively). This process relies on elector honesty in their personal assessments, but it provides a mechanism for addressing differential information on qualifications and expertise. In doing so, it can mitigate cluster voting effects that arise in visible colleges where a group of colleagues vote en masse in support of a candidate colleague irrespective of their personal expertise in the candidate's field.

Since this system involves honesty in elector responses, it opens up new opportunities to game the system. We address this problem by monitoring elector responses and by introducing the prospect of auditing.

Measures to Mitigate Gaming of the System

Distortions arise in any electoral system from voters seeking to achieve certain ends like the promotion of individual fields, institutions, or invisible colleges. The system of peer evaluation leading to an assessment score like X_b offers a new way of gaming the system. For instance, a voter who wanted to promote a particular field (A, say) could return a distended positive evaluation, such as $\varphi = 0.99999$, for candidates in field A, and a distended low evaluation $\varphi = 0.00001$ for candidates outside of A, thereby exacerbating field divisions and avoiding exclusion as an extreme assessment in $\{0, 1\}$. Similar gaming techniques might be applied to evaluating candidates on criteria other than field.

There are various ways to mitigate such distortions. All of these require some monitoring of elector returns. The idea we suggest here relies on auditing (or at least the threat of auditing). It relates to the editorial process by which articles are peer reviewed for publication. The common system of editorial review involves independent referee evaluations which are subsequently assessed by an associate editor or co-editor prior to a final editorial decision. Think of the electors as the referees and the monitors as the senior editorial team of a journal. Peer evaluation for fellowship election can be managed along these lines in the following steps.

(i) A monitoring committee (MC) is appointed with a small group of members representating the main subfields of the discipline. Appointment to the MC may be made by the president on the advice of the society council or executive committee. A new MC is appointed annually for each fellowship election, helping to avoid problems of persistent orientation in decision making. An overview committee comprising the president, vice president, and past president could also selectively audit the MC recommendations, resolving potential "*quis custodiet ipsos custodes*" problems.

(ii) In the ballot, each elector j is requested to return an honest self-selection rating (R_{jia} in the notation above) with regard to each candidate i for whom they return a quantitative peer evaluation. For example, elector j might return $\varphi_{ji} = 0.95$ and R_{ji1}, indicating the elector is a self-regarded expert in the field of candidate i and gives a peer evaluation score of 95% to this candidate. The MC evaluates the responses (or a selection of responses) and if the self-ranking R_{ji1} for high expertise regarding candidate i's research withstands scrutiny (e.g., candidate i has published in the same field as the elector and has an accompanying statement of some authority on the candidate's contribution—see (iii) below), then the peer evaluation $f_{ji} = w_{i1} \times \varphi_{ji} = w_{i1} \times 0.95 = 0.95$ stands (subject to a further possible extreme evaluation trigger check—see (iii) below). If the MC disagrees with the self-ranking, then the peer evaluation is re-weighted as $f_{ji} = w_{ib} \times \varphi_{ji} < 0.95$ with a new weight w_{ib} that is agreed by the MC. If no agreement in the MC is reached on the new weighting value, then φ_i is simply replaced by $\varphi_{i,\text{nom}} = \frac{1}{\#(S_{i,\text{nom}})} \sum_{k \in S_{i,\text{nom}}} \varphi_{ik}$ or $\varphi_{i,\text{valid}} = \frac{1}{\#(S_{i,\text{valid}})} \sum_{k \in S_{i,\text{valid}}} \varphi_{ik}$, as per the earlier discussion. (Here $S_{i,\text{nom}}$ and $S_{i,\text{valid}}$ are the sets of electors who nominated candidate i and who validly—i.e., passed the MC audit—voted for candidate i.) The MC audit may be viewed in the same light as secondary examining and external examining procedures, such as those in the U.K.

(iii) Extreme peer evaluations for $\varphi \notin (L_\varphi, U_\varphi)$ are audited by the MC. But the audit trigger parameters $(0 < L_\varphi < U_\varphi < 1)$ are not announced in the election, so that electors cannot deliberately avoid an audit simply by setting $\varphi = L_\varphi + \varepsilon$ for a candidate they oppose or $\varphi = U_\varphi - \varepsilon$ for a candidate they want elected, with some small $\varepsilon > 0$ intentionally selected to avoid the audit. As with accounting audits, the parameters (L_φ, U_φ) may be individual elector specific and they might be determined by an automated rule that operates outside these parametric settings. For instance, an audit might be triggered if an elector returns a common evaluation of 0.90 for all candidates even though $0.90 \in (L_\varphi, U_\varphi)$. In the audit itself, the MC will look at the self ranking and supporting statements in the voter's electronic return. If the supporting statements on the strength or weakness of the candidates hold up to examination, then the peer evaluation φ_{ji} stands. If not, then the MC will reweight the assessment or simply replace φ_{ji} by $\varphi_{i,\text{nom}}$ or $\varphi_{i,\text{valid}}$ as in (ii). With this system, if an elector is strongly in favor of a candidate and has convincing supporting evidence, then returns of $\varphi = 1$ may well be validated by the MC.

To operate this system, electronic returns require electors to do the following: (a) return a vote in favor of each candidate whom they support (simple approval voting); (b) provide a self selection ranking R_{ji} that reports their level of expertise for each candidate for whom they vote; and (c), if they choose to do so, write a supporting paragraph (less than 200 words, say) justifying their support for the candidate and their subjective assessment of the candidate. Electors are advised that extreme peer evaluations are audited and are therefore encouraged to submit a supporting statement for all candidates for whom they vote and particularly those for whom they record an extreme peer evaluation. Lazy or uninformed electors can simply return their vote and provide no further information, in which case the peer evaluation will be deemed absent (invalid) and automatically set to $\varphi_{ji} = \varphi_{i,\text{nom}}$ or $\varphi_{i,\text{valid}}$. Thus, demands on electors in this system are only greater than simple approval voting when electors individually choose to add further information. The incentive to do so comes from the aspiration that, in addition to their vote, their other input will be taken into account in the election and may contribute materially to the outcome.

With monitoring and auditing in effect the revised version of (1) has the following form. As above, let φ_{ji} be the reported subjective assessment of candidate i by elector j and R_{jia} be elector j's personal ranking of his expertise with regard to candidate i. Let the MC and audit results be represented by the events

$$A_{ji} = \left\{\text{MC validates elector information } \left(\varphi_{ji}, R_{jia}\right)\right\},$$

$$B_{ji} = \left\{\begin{array}{c}\text{MC rejects elector information } \left(\varphi_{ji}, R_{jia}\right) \text{ and} \\ \text{resets } R_{jia} \text{ to } R_{jib} \text{ with new weighting } w_{ib}\end{array}\right\},$$

$$C_{ji}^{r} = \left\{\begin{array}{c}\text{MC rejects elector information } \left(\varphi_{ji}, R_{jia}\right) \text{ and} \\ \text{resets } \varphi_{ji} \text{ to } \varphi_{r} \text{ where } \varphi_1 = \varphi_{i,\text{nom}},\ \varphi_2 = \varphi_{i,\text{valid}}\end{array}\right\}.$$

With these modifications after MC evaluation, the final subjective assessment factor $f_{ji} \in [0, 1]$ of candidate i ascribed to elector j becomes

$$f_{ji} = \sum_{a=1}^{A} w_{ia} \varphi_{ji} \mathbf{1}_{R_{jia}} \mathbf{1}_{A_{ji}} + w_{ib} \varphi_{ji} \mathbf{1}_{B_{ji}} + \sum_{r=1}^{2} \varphi_r \mathbf{1}_{C_{ji}^{r}}, \tag{3}$$

where A is the number of weights $\{w_{ia} : a = 1, ..., A\}$ and $\mathbf{1}_H$ is the indicator of H.

Collating the Return Information

Having determined each voter's f_{ji} by this process, the aggregate component $f_{(i)} = \frac{1}{\#(S_{\text{all}})} \sum_{j \in S_{\text{all}}} f_{ji}$ represents the average subjective view of the voting electorate on candidate i. The subjective contribution to the overall accomplishment factor X_i of candidate i is then $X_{bi} = \frac{1}{2} \times f_{(i)}$, whose support is $U_b = \left[0, \frac{1}{2}\right]$. So X_{bi} carries an implicit weight of $\frac{1}{2}$ in X_i. Importantly, X_{bi} places demands on individual electors that go beyond simple Yes/No or rank voting schemes. Each elector must translate a subjective judgment of a candidate into a quantitative subjective score for that candidate, indicate the elector's own expertise in the research field of the candidate, and provide a statement that supports the subjective assessment. If the elector wants to be confident the returned score will count in the election, then the elector must be prepared for an audit. If the return is audited and fails, then the elector's peer evaluation will be reweighted in the audit or replaced by a community based evaluation.[10]

The elector retains voting privileges to vote on the candidate. This vote and the subjective assessment end up playing dual roles in the election. Thus, voters influence the election of each candidate by transporting their personal information and subjective assessment of a candidate into a score that affects the merit threshold of the candidate as well as by a direct Yes/No vote on the candidate. Gaming is discouraged by the threat of an audit by the MC.

Based on these two components, the overall accomplishment factor is computed as $X = X_a + X_b$ (omitting the candidate index i). Obvious modifications involve differential weights for the objective and subjective elements X_a and X_b in the scheme, with corresponding differences in the supports U_a and U_b. For example, we might set $X_a = \lambda \times \frac{N}{M} \in [0, \lambda]$ and $X_b = (1 - \lambda) \times f \in [0, 1 - \lambda]$ for some preassigned weight $\lambda \in [0, 1]$. Then, when $\lambda = 0$ (respectively, 1) only subjective (objective) assessments are taken into account.

In order to control the influence of the additional information embodied in X on electorate voting, parameters may be set to determine upper (τ_U) and lower (τ_L) merit thresholds for election. Thus, τ_U defines the (upper level) proportion of votes that is required for election when additional information X takes some minimal value ($\gamma_\ell \geq 0$). Similarly, τ_L defines the (lower level) proportion of votes that is required for election when additional information X takes some maximal value ($\gamma_u \leq 1$).

With these settings and given the additional information X, the formula for the merit threshold has the form

$$\tau = \tau_U 1_{\{X < \gamma_\ell\}} + \tau_L 1_{\{X > \gamma_u\}} + \left[\tau_L + (\tau_U - \tau_L) \left\{ \frac{\gamma_u - X}{\gamma_u - \gamma_\ell} \right\} \right] \mathbf{1}_{\{\gamma_\ell \leq X \leq \gamma_u\}}. \tag{4}$$

[10]The process is analogous to the peer review process in academic journals, where referee reports and recommendations are evaluated by an associate editor or co-editor and editor before an editorial decision is made. Referees often voluntarily return in their cover letters an indication of their own expertise in the field, rather like R_{jia}.

For each candidate (i) in the election, the corresponding merit threshold τ_i is computed using formula (4) together with the component information $X_i = X_{ai} + X_{bi}$ for that individual. The decision rule in the election of candidate i then depends on the actual voting percentage (V_i) supporting that candidate in the election. If $V_i \geq \tau_i$ so that the percentage of actual votes meets or exceeds the candidate's merit threshold (τ_i), then the candidate is elected. Symbolically, $E_i = \mathbf{1}_{\{V_i \geq \tau_i\}}$ gives the election outcome ($1 =$ success; $0 =$ failure).

In practice, the main effect of (4) is to require a higher percentage of votes in the election for candidates with less demonstrated accomplishment as measured by X. Peer support in the election votes must then be decisive to outweigh the effect of less demonstrated accomplishment. When the merit threshold bound parameters are equal, i.e., $\tau_L = \tau_U = \tau$, the datum X has no effect on the outcome which is then determined solely by some specified threshold level for election (τ), as commonly occurs in current societal practice (e.g., in fellowship elections of the Royal Society and the Econometric Society—see Section A.4 of the Appendix).

Illustration

To clarify the workings of the above formula, we may take a specific parametric form with $\tau_U = 0.5, \tau_L = 0.2, \gamma_\ell = 0.25$, and $\gamma_u = 0.75$. The merit threshold for a candidate (again omitting the index i) then has the following explicit form:

$$\tau = \begin{cases} 50\% & \text{if } X = X_a + X_b < \frac{1}{4} \\ 20\% + 30\% \left\{1 - 2\left(X - \frac{1}{4}\right)\right\} & \text{if } \frac{1}{4} \leq X \leq \frac{3}{4} \\ 20\% & \text{if } X > \frac{3}{4} \end{cases},$$

where the following situations hold:

(i) $X_a = \frac{1}{2} \times \frac{N}{50}$, where $N = \min(2n_1 + n_2, 50)$ with $n_1 =$ number of sole authored publications in core designated journals and $n_2 =$ number of co-authored publications in core designated journals;

(ii) $X_b = \frac{1}{2} \times f$ and $f = \frac{1}{\#(S_{\text{all}})} \sum_{j \in S_{\text{all}}} f_j \in [0, 1]$ with f_j (more precisely f_{ji} for candidate i) determined as in (3).

In this example, candidates with an accomplishment factor X that is lower than $\frac{1}{4}$ must receive 50% or more votes in the election to be elected. Likewise, candidates with an accomplishment factor that exceeds $\frac{3}{4}$ need only receive 20% or more votes in the election to be elected. In this manner, quantitative evidence on accomplishment and collective peer evaluation influence the election outcome by adjustment of the election threshold, reducing requirements for candidates who have and are perceived to have a strong track record in the discipline. The parameter settings $\{\tau_U = 0.5, \tau_L = 0.2, \gamma_\ell = 0.25, \gamma_u = 0.75\}$ are illustrative. Some computations that show the effect of changes in these parameters, and those that determine the density of X and the implied density of τ are reported in the following section.

One likely effect of the introduction of evidence-based merit thresholds is a reduction of the distortion bias that can arise from cluster voting for less (materially) qualified candidates. As mentioned earlier, pre-eminent institutions often have many existing society fellows and the electoral strength of these voters can be decisive in securing election for colleagues who may be less materially well qualified than others at less eminent institutions. The presence of such candidates at pre-eminent institutions might itself be regarded as an endogenous indicator of quality and may therefore, in some formulae, enter into the merit threshold calculation—for example, in the case of the mechanism described above, it may enter through the peer review factor X_b by way of the individual quality assessment φ_j. However, we can expect that to be elected when an evidence-based merit threshold is used, such candidates will generally require a greater percentage of the votes cast in the election if their quantitative merit score X is below the threshold γ_ℓ.

Another mitigating effect in the use of an evidence-based merit threshold is the reduction of bias arising from invisible college coalition voting for candidates within certain fields. In such cases, electors may vote in coalition for some candidates, making it easier for those candidates to reach a predetermined fixed threshold of votes. Under (4), however, the peer view of the entire electorate is taken into account in the measurement of X_b, peer evaluation of the strongest supporters may be audited, and the track record of material accomplishment of the candidate comes into play in determining X_a. These factors end up determining the merit threshold that is needed for a candidate's election and this broad basis of extra information on the candidate will tend to dilute the impact of coalition voting in the election.

4. THE MERIT THRESHOLD DISTRIBUTION

Some implications of the above formulae can be explored by determining the exact distribution of X and the implied distribution of τ. The latter is the main focus and reveals how various degrees of component information affect the perception of merit and drive the threshold level.

To proceed, we need to make some further assumptions. It is convenient to assume that the electorate population is large enough for the key components to be continuously distributed, leading to a distribution of X over the interval $[0,1]$. The resulting distribution of τ has a mixed continuous and discrete form comprising a double spike and a smooth distribution. There are point masses at the upper and lower threshold levels τ_U and τ_L, and a continuous distribution applies between these thresholds. In particular, if $p_X(x)$ is the density of X on its support $[0,1]$ and $p_\tau(t)$ is the density of τ over (τ_L, τ_U), the upper and lower threshold probabilities are given by

$$p_U = P(\tau \geq \tau_U) = \int_0^{\gamma_\ell} p_X(x)\,dx, \quad p_L = P(\tau \leq \tau_L) = \int_{\gamma_u}^1 p_X(x)\,dx, \tag{5}$$

and the density by

$$p_\tau(t) = p_X\left(\gamma_\ell + \frac{\gamma_u - \gamma_\ell}{\tau_U - \tau_L}(\tau_U - t)\right)\frac{\gamma_u - \gamma_\ell}{\tau_U - \tau_L}, \quad \text{for } t \in (\tau_L, \tau_U). \tag{6}$$

The distribution of $X = X_a + X_b$ is a convolution of its two components. The objective component

$$X_a = \frac{1}{2} \times \frac{N}{M} = \frac{1}{2} \times \frac{\min(Y, M)}{M} \in \left[0, \frac{1}{2}\right]$$

has a probability mass at $\frac{1}{2}$ arising from the upper bound M on admissable publications data. As remarked above, this bound delimits quantity effects in bibliometric data to a preassigned level M. It follows that the density $p_a(x)$ of X_a will in general have a spike at the upper bound $\frac{1}{2}$. The subjective component X_b has density $p_b(x) = 2p_f(2x)$ where p_f is the density of $f \in [0, 1]$. If $p_{a,b}(x_a, x_b)$ is the joint density of (X_a, X_b), then the distribution of X has the convolution form

$$p_X(x) = \int_0^{x \wedge 1/2} p_{a,b}(x - t, t)\,dt.$$

With this structure it is possible to obtain the exact density $p_X(x)$ in terms of the density[11] $p_Y(y)$ of Y and the density $p_b(x)$ of X_b. Derivations are given in Section 7.1 of the Appendix. Under the simplifying assumption of independence[12] between X_a and X_b, it is shown there that

$$p_X(x) = 2M\int_0^{x \wedge 1/2} p_Y(2M(x - t))\,p_b(t)\,dt + \mu_M \times p_b\left(x - \frac{1}{2}\right) \times 1_{\{x \geq \frac{1}{2}\}}. \tag{7}$$

The density $p_X(x)$ can have a jump at $x = \frac{1}{2}$. The size of the jump depends on the parameter $\mu_M = P(X_a > M)$ corresponding to the probability that an individual's

[11]In working out the exact distribution theory, it is convenient to let the aggregate publication component $Y = 2n_1 + n_2$ have a continuous density. A corresponding discrete equivalent can be computed by integration over cells of unit length covering the integers.

[12]Peer evaluation, as embodied in X_b, is surely influenced by accomplishment, as measured in X_a. So we can expect X_a and X_b to be dependent in general. In a large elector population, however, there may be many electors who have little knowledge of the accomplishment of candidates outside their fields, in which case the assumption of independence may not be fatally violated. See the discussion in Section 7 on the Econometric Society fellowship elections, where the Council of the Society noted that "works of several well known nominees had been read by only a few Fellows" and "ballots show that some nominees failed of election primarily because their work was to a large extent unknown to the Fellows."

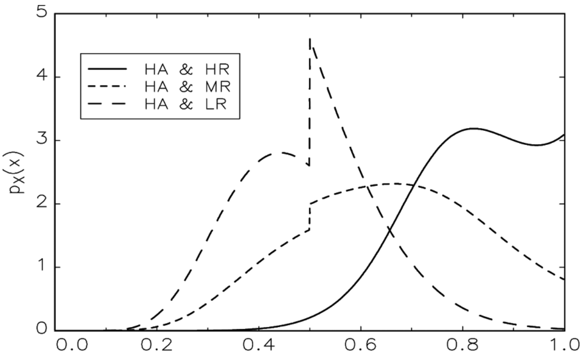

FIGURE 1 Densities of $X = X_a + X_b$ for distributions of X_a and X_b corresponding to high accomplishment (HA) and high peer review (HR), mixed peer review (MR), and low peer review (LR).

publication count exceeds the designated count M. The size of the jump also depends on the value of the density $p_b(y)$ of X_b at $y = 0$ and is zero when $p_b(0) = 0$ so that $p_X(x)$ is continuous in that case. Thus, the population of candidates with a publication count in excess of the designated maximum produces a point mass in the distribution of X_a giving a spike and smooth density $p_a(x)$ of X_a and, upon convolution, the spike can translate into a jump in the density of X at $x = \frac{1}{2}$, the upper point of the domain of X_a, when $p_b(0) > 0$.

Fig. 1 illustrates these possibilities for various accomplishment and peer review distributions that are fully described later in the Appendix of the article. Discontinuities in the density $p_X(x)$ of X typically arise when there is conflict between objective evidence as it is embodied in the distribution of X_a with a point mass at level M, and peer review opinion when this produces a positive density to X_b at zero. In Fig. 1, the two discontinuous densities shown in the broken lines of the figures arise when high material accomplishment (manifest in the X_a distribution with $P(X_a \geq M) > 0$) couples with subjective peer review that includes some strong negative opinion (associated with a density $p_b(x)$ of X_b for which $p_b(0) > 0$, reflecting a cluster of peer opinion around zero). The probability mass in X_a leads to a jump in the density at $X = \frac{1}{2}$ and the negative peer review effect leads to a decline in the subsequent density of X as X approaches its upper limit of unity. The stronger the negative peer review, the sharper the ultimate decline in the density, as manifest in the high accomplishment with low peer review (HA and LR) case in the figure.

This simple distributional exercise shows how accomplishment and peer review interact to produce a wide variety of possible informational densities on a candidate that affect election probability. More extensive illustrations and detailed computations for selected parameter configurations are given in the Appendix.

5. IMPLEMENTATION

Formula (4) seeks to bring both quantitative and qualitative information to bear on fellowship elections. The intent is to ensure that substantive research accomplishments and collective peer review count in electing new fellows, so that the threshold of support from existing fellows is greater (at most $\tau_U \times 100\%$) for individuals with fewer accomplishments and is less (at least $\tau_L \times 100\%$) for those with greater accomplishments. While the publication count component X_a is primarily quantitative, this measure also has an implicit qualitative element by virtue of the journal selection and the peer judgments that underlie publication. Journal selection can obviously be modified by a society's governing council to reflect changing standards and evolution of the discipline as it manifests in core journals. The component X_b allows for the full voting electorate to return subjective assessments of the candidate. These assessments offer the opportunity to take account of a wider set of qualifications (such as acknowledged impact of research on other disciplines, outstanding pedagogical work, mentorship, and contributions to software development) so as to more fully reflect the professional contributions of an individual candidate for election. The measure X_b then reflects the overall peer assessment of the candidate across the voting electorate.

Implementation of this procedure requires parameter inputs, data collection and some computation. The process itself can be coordinated by a society's secretary and is readily accomplished online using a web server. Senior officers of the society would need to take a leadership role in deciding on the most suitable version of the system, including the form of the monitoring committee, appointment of an election officer to oversee the election, and the audit process to overview subjective evaluation of candidates and self ranking by electors.

The key steps are detailed below.

1. **Prior Parameter Settings.** Parameters that appear in formula (4) need to be set by the society, presumably through its governing body or council. The parameters that require prior setting are as follows:

    (a) The domain parameters $\tau_L, \tau_U, \gamma_\ell$, and γ_u that appear directly in formula (4);
    (b) The bound parameter M that specifies the upper bound on the number of publications (or other bibliometric information) considered in the quantity measure X_a;
    (c) The relativity parameter ρ (currently 2 in (4)) that distinguishes sole authored from co-authored publications in the publication count $Y = 2n_1 + n_2$;
    (d) The weight parameter $\lambda \in [0, 1]$ (currently $\lambda = \frac{1}{2}$ in (4)) which allocates a weight of λ to quantitative information X_a and a weight of $1 - \lambda$ to subjective assessment X_b;
    (e) The audit trigger parameters $\left(L_\varphi, U_\varphi\right)$ and any additional audit triggers (such as zero variance assessments or self-rankings);

(f) The maximum length of the supporting statements in elector evaluations of candidates (such as 200 words).

2. **Nominations.** Candidates for fellowship need to be nominated by those members enfranchised to vote in fellowship elections. The information required in a nomination typically would include the following kinds:

   (a) A nominating statement of some designated length (such as 200 words);
   (b) A list of n^* of the candidate's most influential publications (such as $n^* \leq 10$);
   (c) Citation data on those n^* publications;
   (d) Summary quantitative information on publications including the pair (n_1, n_2) of sole authored and co-authored publications;
   (e) A subjective assessment factor $\varphi \in [0, 1]$ of the candidate by the nominator.

3. **Deadlines.** A deadline is set for the submission of all nominations, including the objective and subjective information that must accompany the nomination. A date for the opening of the election is set and a deadline for the return of ballots.
4. **Criteria for Nomination.** A criterion must be set to determine those nominations that will be taken to the electorate for voting. For example, all nominated candidates might be submitted to the electorate or only those candidates who have received at least a certain number $n^\#$ of supporting nominations (typically $n^\# \geq 3$). In the case of a candidate receiving supporting nominations, each nominator may return a different φ_i.
5. **Ballot Information.** Electors receive online ballot information that includes a list of nominees, the nominating statement for each candidate (i) with a list of the supporting nominators (j), the objective component X_a, and the subjective component $X_b^{\text{nom}} = \frac{1}{2} \times f^{\text{nom}}$, where $f^{\text{nom}} = \frac{1}{\#(S_{\text{nom}})} \sum_{k \in S_{\text{nom}}} \varphi_k$ is the average subjective assessment factor of the candidate from the nominating electors, which is calculated from the assessments φ_k submitted by the nominating electors. As discussed in Section 7, dropdown menus in the online ballot might be available to display further bibliometric and citation information on the candidates.
6. **Elector Returns.** The elector returns include as follows:[13]

   (a) The vote itself (Yes/No);
   (b) The elector's subjective assessment factor for each nominated candidate (voluntary);
   (c) The elector's self rankings of their expertise in the field(s) of each candidate (voluntary);

[13]As indicated, items b, c, and d are voluntary. Where the information is not provided a default assessment is created as discussed in Section 3.

(d) Statements by the elector in support of the elector's own peer evaluation of each candidate (voluntary).[14]

7. **Vote Counting.** After the election, votes for each candidate are counted. The subjective assessment data submitted in the election returns by each voter are aggregated to produce the subjective component X_b and combined with the objective data X_a to produce X and compute the merit threshold τ for each nominated candidate. Actual votes for candidate i are expressed as a percentage (V_i) of all valid votes cast and compared with the candidate's merit threshold τ_i. Candidate i is elected if $V_i \geq \tau_i$.

Some computations that are given in the Appendix offer general guidance on the impact of different parameter settings. More explicit evaluation that is relevant to a particular society can be conducted through simulations that mirror ingredients within the formula (such as particular parameter settings) that produce outcomes like those of earlier societal elections (conducted without the formula) and alternative outcomes that result from other parameter settings (as counterfactual tests to evaluate the sensitivity of outcomes to parameter changes). These simulations and guidelines can be assessed by the governing body to determine the adequacy of certain parameter ranges for society purposes.

Importantly, even without formulae such as (4), elections require some parameter settings, which are then implicit in the simpler system. For instance, in an election system where only votes count, the merit threshold τ for election must still be determined. Such a system has $\tau = \tau_L = \tau_U$, and information in X is then completely ignored.

Any quantitative information like (n_1, n_2) that is submitted in the nomination can be cross-checked through an online service that provides automated harvesting of publication data. To ensure uniform treatment across candidates, a society may require that all publication data be obtained (and checked) in this way from a reputable bibliometric harvester, much as some universities now do in conducting research assessment exercises on faculty and departments.

6. BROADER ISSUES

Beyond fellowship elections, learned societies and academia lie much wider qualitative assessment issues facing humanity. Rankings of "excellence" and "quality" are widespread in modern society touching every aspect of life—the environment, health care, crime, politics, education, finance, and the economy. Economists are far from being the only

[14]A single sentence may be sufficient, such as "I know the candidate's work well, can confirm that the papers listed in the nominating statement are significant contributions, and am strongly supportive of election," or "I have read several of the candidate's papers, find the research to be incremental and believe it is too early to support election of this candidate."

professionals concerned about the rapid advance in the use of data from bibliometric and other online sources in such quality assessments. The President of the Institute of Mathematical Statistics recently exhorted the broad community of statisticians on these matters, saying that

> ... almost alone we have the skills to respond to them, for example by developing new methodologies or by pointing out that existing approaches are challenged. To illustrate the fact that issues that are obvious to statisticians are often ignored in bibliometric analysis, I mention that many proponents of impact factors, and other aspects of citation analysis, have little concept of the problems caused by averaging heavy tailed data. We should definitely take a greater interest in this area. (Hall, 2011).

Some of the underlying problems in this area relate to the issue of index selection for latent concepts like "academic excellence" (as we have been discussing here), "quality of life" or "well being," all of which rely on multivariate data, multi-indicators, and subjective assessment. Index number construction has a long history in economics going back in modern form for over one and a half centuries, whereas interest in multidimensioned indices for concepts such as deprivation or wellbeing is more recent (e.g., Atkinson and Bourguignon, 1982; Atkinson, 2003; Maasoumi, 1999). Early investigators were challenged by the paucity of data. But in modern society, terabytes of information descend weekly on libraries, statistical agencies, social networks, international agencies, and central banks.[15] Sorting, storing, and sensibly describing these vast quantities of data are major undertakings even for the academic rankings that have proliferated in recent years (e.g., Baltagi, 2007). But analyzing and creating useful indices from very high dimensional electronic data sets is a far greater challenge that requires novel methods and algorithms from the rapidly expanding field of high dimensional statistical analysis while also accounting for possible effects from averaging heavy tailed data (Hall and Miller, 2010), as indicated in the above citation.

At the same time as electronic information has mushroomed in society, there has been growing disquiet over inadequate uses or analyses of such information and the possible consequences of misleading information. A high profile example that was catapaulted into public awareness by the global financial crisis (GFC) is the ranking of financial instruments by credit rating agencies. During the GFC massive losses were incurred in the market for highly structured financial instruments such as collateralized debt obligations (CDOs) despite the AAA and Aaa ratings assigned to the instruments by these agencies. The elevated point ratings of CDOs did not reflect the concentration in risk and associated

[15] Wikipedia reports that the U.S. Library of Congress adds 5 terabytes (TB) of information per month, that internet traffic is estimated to be around 160 TB a second, that Germany's climate research centre generates 10,000 TB of data per year, that Wikipedia itself had a 5.87 TB dump of raw data in January 2010, and that Google processes about 24 petabytes of user-generated data per day.

vulnerabilities arising from the bundling of very large numbers of similar securities. Ensuing criticisms of rating agency practices have pointed to the bias in their risk assessments of highly structured financial products, but these criticisms have led to little change in analysis and reporting. To address these criticisms, greater transparency about the dimensional composition of financial products is required, as well as more thorough investigation and data analysis, coupled with improved reporting mechanisms that more accurately reflect essential product features and associated risks. Such improved reporting requires much more than a one-dimensional ranking involving a simple letter grade.

In addition to the task of trawling through vast data sets to produce quantitative information that is relevant to a quality assessment exercise, we must also accommodate subjective human assessments in quality indices of human existence. Just as peer review is considered an essential element in academic merit assessment, human judgment is an important factor in most other areas of quality assessment. For example, in ranking city quality of living, published indices normally rely on neutral objective information about a host of measurable elements such as air quality, crime rates, infrastructure, and amenities. But city residents have their own personal assessments of cultural and recreational facilities, a city's ethnic diversity, its climate and environment, and its amenities. Combining these dual sources of objective and subjective data, accounting for the differential background information that is inevitably involved in human assessment, and understanding the statistical implications of the averaging process involved in index construction from high-dimensional sources of data is the much broader arena within which the task of meritocracy assessment resides.

7. ECONOMETRIC SOCIETY FELLOWSHIP ELECTIONS

While many of the issues we have raised have broader import in the general area of quality assessment, they are particularly relevant to the Econometric Society (ES) which has been an honor society with a meritocracy of fellows from its inception. Since 1960, the ES has operated an approval voting system for new fellows whereby new candidates for fellowship are elected if the fraction of voters who approve the candidate is greater than a certain threshold (now 30%). Setting the proposals of the present article in the context of the ES highlights some of the advantages of an extended system of election with greater information content but also reveals some practical difficulties of implementation. We consider these briefly here and encourage a wider discussion to promote improvements in the current system of election. Some aspects of the evolution of ES election procedures are also of interest and will be discussed in what follows.

1. Since the ES now publishes three journals, a simple starting point in the construction of an objective quality measure such as X_a would be to use these journals as the core designated journals and count each candidate's publications in these journals, discounting multiple authored papers by some scale factor, as in the measure Y. While

some members of the ES may see this system as reasonable, others may view the exclusion of leading field journals or other general interest journals as unfair to those fields that are not well represented (or increasingly under-represented) in the society journals (c.f. the remarks in footnote 2). After all, it took the society 80 years to expand its own journal offerings,[16] whilst the subject reach of quantitative economics, economic theory, and econometrics grew massively producing a vast offering of leading field journals and related general interest journals in which the ES played no role. There is no immediate way to reconcile these different perspectives, just as it took the ES a long time to respond to calls for the expansion of its journal offerings. But active consultation with members does help. For instance, society members (or fellows) could be asked to determine (or vote on) the issue themselves by a process of core journal augmentation. Each year members could nominate new journals for inclusion in the core and in an annual election journals could be added (or eliminated) by an approval voting system with some designated threshold of votes for inclusion. Such a system would accommodate subject-based evolution that is reflected in the growing importance of certain journal outlets outside of the ES offerings. As in the current paper's analysis, such a voting system could be modified to take into account more information than simple Yes/No voting.

2. Evidence mentioned in the introduction in footnote 3 indicates that for some time the ES has not been electing many potential fellows who rank highly in terms of various quality measures, including multiple publication counts in the society's flagship journal *Econometrica*, while others are elected who have no publications in *Econometrica*. Part of this information about candidate quality is already present in nominating statements that are available to electors prior to voting. So, a natural question is why voters are apparently not taking account of this information, given actual election outcomes. One possible explanation is that electors in the relevant subfield of these candidates are acting on the information but lack the voting power to meet the required threshold. The widely recognized difficulty in electing candidates in econometrics, particularly econometric theory, supports this explanation. Exceptions where election is easier in such subfields typically occur for candidates in top tier institutions in North America, where voting power is heavily concentrated giving those candidates a comparative advantage because of additional collegial cross-field voting in support of their election. Another possible explanation is that many voters do not use the information in the nominating statements and only vote for the candidates that they know well. Or they may simply find it too time consuming to research all the candidates and make quality comparisons. In such cases, electors may need to be provided with key summary

[16]Proposals to expand the list of ES journals go back a long way. For example, to reflect the substantial expansion and growth in diversity in the discipline by the 1980s, the author proposed to the society in 1982 that *Econometrica* be expanded into two flagship journals, one subtitled in *Economic Theory*, and the other in *Econometrics*, much as the Institute of Mathematical Statistics created the *Annals of Probability* and *Annals of Statistics* out of the original *Annals of Mathematical Statistics*.

information that includes quality indicators on each candidate directly online in the ballot form. This information can be delivered in the ballot form itself or via a pull down menu which provides ready access to the data and more sophisticated options, like selecting for perusal only those candidates with at least some (user-specified) number of publications in certain journals. Voters might also be required to check some such information on every candidate before their voting option is enabled, thereby enforcing exposure to some relevant data on the candidates.[17] These options conform with the suggestions of the present paper to increase the information content available to electors on the ballot, to enforce some exposure to the data, and to incentivize a greater level of participation by electors in the election process.

3. Field under-representation could also be partly addressed by implementing a version of a quota system, like that operated by the Royal Society and the National Academy of Sciences, or a section system like that operated by the British Academy—see Appendix A.4 for further information about the election processes in these societies. Quota upper limits placed on the number of candidates elected in a particular field or section help to ensure representation in each of those sections, much like a proportional representation system in political elections.
4. Allowing for an abstention option in approval voting would provide a simple way for electors to register insufficient knowledge of a candidate to cast a vote. Abstentions would be excluded from the electoral count proportion and, as a referee suggested, may therefore help to reduce bias against candidates whose work is less known. This facility might be supplemented with a requirement that a minimum number (or proportion) of electors cast votes to eliminate extreme outcomes such as a 50% vote in favor of a candidate based on only two people voting.
5. A further mechanism to enhance elector participation, suggested by a referee, is for voters to record their own suggested threshold percentage in their ballots. The median response might then be used to set the threshold voting percentage. This has the advantage of simplicity and it endogenizes the threshold according to voter preferences, instead of by council mandate. On the other hand, such a threshold is global and not individual specific. A variant of this method might take account of the distribution to set upper and lower bounds on the threshold that might be used in conjunction with quantitative and qualitative assessment data.

These are just a few of the many possibilities that might be considered in making the ES ballot system and its fellowship more representative of the discipline's immense and growing diversity. All of these possibilities involve increased use of data and/or elector participation in the election process.

In comparison with the procedures outlined in this article, the present ES nomination and voting arrangements are simplistic and the ES fellows ballot requires no input from

[17] At present, there is no requirement in the online system for electors to look at the nominating statement before voting.

electors other than an approval vote. The original procedures of the ES in selecting new fellows were substantially more complex than the present arrangements. They involved discussion and debate about potential candidates within the Council (the body charged with making nominations) prior to actual nomination and required far greater input from electors. Readers are referred to Bjerkholt (1998) for a history of the evolving processes of ES fellows elections during the 1930s, the role of Ragnar Frisch in these developments,[18] and some of the controversies that arose during this period. Importantly, some guidelines for qualifications were laid down by the Council, requiring candidates to be an economist, statistician, have some knowledge of higher mathematics, and have made some original contributions to knowledge, including economic theory.[19,20] Demands were also placed on electors that required them to acknowledge whether or not they had read any works by the candidate and whether they had "critically scrutinized representative works of nominees" (Bjerkholt, 1998, p. 54). The electors (who by 1935 were the ES Fellows) were asked to categorize candidates according to the electors' subjective assessments of candidate distinction and record electoral votes into four classified columns (detailed in Bjerkholt, 1998, p. 54) as follows:

"(a) Votes for nominees considered to have at least as high qualifications for fellowship as the average of the present Fellows";

"(b) Votes for the better half for those already marked in colum (a)";

"(c) Votes for those nominees in column (b) who it is *very strongly* felt should be elected";

"(d) Votes against those nominees whose election is *very strongly* opposed."

[18]For further reading on Frisch, see Bjerkholt (1995, 2005, 2015).

[19]Correspondence between Irving Fisher (President of the ES) and Ragnar Frisch (Editor of *Econometrica*) provided guideline criteria for the selection of Fellows. In a letter to Council on 12/01/1932, Fisher appended a statement listing 5 qualification categories for selecting fellows: "1. They should be economists; 2. They should be mathematical; 3. They should be statisticians; 4. They should have made some original contributions; 5. Some of these contributions should be in economic theory." In a response to Fisher on 1/11/1933, Frisch indicated agreement on these criteria, which he re-articulated as: "1. The candidate must be an economist acquainted with economic theory; 2. He must have a mathematical foundation; 3. He must have some knowledge of statistics; 4. He must have done some original work; 5. Some of this original work must have been in economic theory." This correspondence is contained in the Frisch correspondence collection of the National Library of Norway. I am obliged to Olav Bjerkholt for sending me copies of this correspondence.

[20]These qualifications for fellowship of the ES are very close to the requirements for membership of the society that were originally laid out in a letter of June 17, 1930 to multiple recipients from Irving Fisher, Charles Roos, and Ragnar Frisch (from the Ragnar Frisch Archive, the National Library of Norway). These requirements limited membership to those who: "(a) are thoroughly familiar with general economic theory; (b) have a working knowledge of mathematics as applied to economic theory and statistics; (c) have some knowledge of accounting; and (d) have published an original contribution to economic theory or to the analysis of such economic statistics or accounting as have a definite bearing on problems in economic theory." The reader is referred to Bjerkholt and Qin (2010) for further discussion of this history, ES membership elegibility, and related matters.

Votes in categories (a), (b), and (c) were assigned respective weights 1,3, and 5, and aggregated to give totals for all candidates. In the 1935 election, four candidates with the highest score were elected. In announcing the election outcome, the Council noted

> A surprising result of the vote was the discovery that works of several well known nominees had been read by only a few Fellows. Indeed, the ballots show that some nominees failed of election primarily because their work was to a large extent unknown to the Fellows. (*Econometrica*, Vol. 3, pp. 477–478; qtd. in Bjerkholt, 1998)

These early fellowship elections in the ES therefore revealed information not only about the candidates but also about existing fellows. In placing greater demands on the electorate, in encouraging the reading of representative works of candidates, and in weighting subjective assessments, these procedures of the original ES fellows elections have elements in common with the ideas outlined in the present work. Online electronic dispersal of nominating materials and ballots and the ready availability of bibliometric data help to facilitate more complex data-intensive procedures and open up an array of new possibilities in terms of information dissemination, assessment mechanisms, and voting protocols that were unavailable in the early years. In the light of these vast changes in our computing and communication facilities, it may seem ironic that the ES nomination, ballot, and voting protocols have become so simplistic, placing virtually no demands on the electorate other than an approval vote. In spite of the simplicity of the electronic voting procedure, nonvoting appears to be a persistent problem in ES fellowship elections.[21] Various measures can be implemented to help improve participation rates, such as greater transparency in the elections (e.g., reporting the proportion of nonparticipants by country or region), leadership from the council and society president urging participation, or use of the regional standing committees to contact fellows in the region and encourage nominations and voting.

8. CONCLUSION

The focus of the present contribution is the appraisal of credentials, the operational use of available quantitative information and the pooling of human judgment across a population of voters in the process of electing new members to a meritocracy. Some of the problems addressed here might also be studied in a dynamic voting environment. A learned society is a social institution in which the size of its fellowship (itself an electorate) is endogenous since it is determined by voting decisions taken by this same electorate over time. Such dynamic voting problems have been studied in the economic theory and behavioral literature, where new complexities have been discovered. In exploring club

[21] As one referee put it, “my impression is that a significant number of fellows do not vote in an election.”

voting decisions, for instance, Roberts (1999) has shown that dynamic voting on club size leads to time inconsistent outcomes and intrinsic steady states in the system that are determined by the voting dynamics. Acemoglu and Robinson (2000) have developed a dynamic model of the voting franchise that seeks to explain gradual processes of reform and democratization such as the emergence of western democracy. More general problems of endogenous social choice and policy determination have been studied recently in Luganoff (2009). This research in economic theory is relevant in the current setting of meritocracy voting because it focuses on the evolution of the voting franchise over time and the effects of this endogeneity on institutional structure and reform. On the other hand, none of this work addresses the issue of appraisal that is fundamental to meritocracy.

The goal of 'measuring merit' is undoubtedly elusive. But as the header to this article entreats, the difficulty of the challenge should not prevent the attempt. Within economics and more broadly among the social sciences, theory and measurement are seen as twin sisters that work in unison to advance our understanding of human behavior and society. It surely befits such disciplines and particularly economics, so often regarded as the queen of the social sciences, to pioneer a way of bringing the "theory quantitative" and "empirical quantitative" into societal decision making on matters as fundamental to a meritocracy as fellowship elections.

The formulae given here are a first step in addressing these issues. The specific rule (4) is designed to assist in the merit selection process by explicitly taking into account subjective assessments of individual candidates for election as well as direct quantitative measures of quality such as publication numbers in learned journals, rankings or citations. As we have argued, quantitative assessment rules may help decision makers widen the effective electorate of opinion, thereby enhancing the information set that is available for consideration in evaluating candidates. Information on publications ends up reflecting assessments and recommendations that are sought in the peer review process. Citations may be interpretated to provide some information about received opinion on research (or its relevance or neglect), while acknowledging the many caveats regarding these interpretations—as discussed, for example, in Cozzens, 1989, and Brooks, 1986.[22] The import in both cases is that a wider body of views and material evidence comes into consideration when the information is embodied in a merit threshold for election.

In this process, the demands on voters and decision makers are greater than in simple Yes/No or rank order voting. But if voters want to avoid these demands and simply record an approval vote, then they may choose to do so. Our view is that the

[22]The interpretation of citation data has spawned a large literature that reveals the sociological complexity of the practice. For example, in her study of the factors involved in citation, Cozzens (1989) notes that the recognition/reward element of citations neglects rhetorical factors such as convenience ("the easier it is to fit the knowledge claims of the article into the rhetoric of later papers, the more likely the article is to be cited"); and Brooks (1986) reports empirical evidence from author surveys covering a wide range of academic departments that 70% of the citations surveyed arose from a "complex interplay of multiple citer motives" involving positive credit, negative credit, and service to the reader elements, some of which may be conflicting.

demands are to be encouraged, especially in a democratic voting system involving highly qualified electors whose information and skill sets are broad and heterogeneous. As we have explained, the demands enable voters to influence the election of candidates in several ways. They report their subjective assessment of a candidate into a numerical score that combines, possibly in a weighted form that takes into account their self selected field expertise, with the judgments of other voters to produce an overall peer review assessment. This component then combines with material information about the candidate to determine the candidate's merit threshold. Electors also record an individual vote on the candidate which combines with other votes to determine the actual voting percentage in favor of the candidate's election. The subjective assessment, the self expressed field expertise, and the individual vote all influence the final outcome. And an audit process helps to ensure elector integrity in election responses.

The formulation given here is a beginning. Obviously a great deal more work can go into its further mathematical development, into the use of voting theory in its design, and into its online implementation. More attention to data sources, bibliometric measures, and the quantification of subjective assessment all seem desirable. Empirical work may also be possible using past fellowship election data to determine parameters implicit in existing rules and to perform counterfactuals. The present article will have achieved its immediate goal if it stimulates further thinking on these issues and on the general problem of quantifying the assessment of merit. The longer term and more ambitious goal of this research is to assist learned societies in the complex task of sensibly utilizing wider information sets in critical decision making on meritocracy elections, so that these elections more accurately mirror research accomplishment in the discipline. Research on this topic is important not only for learned societies but for the many other instances in academic life where merit assessment is a critical matter in the careers and lives of our colleagues.

A. APPENDIX

A.1. The Merit Threshold Distribution

To find the distribution of τ, we need the distribution of X, which in turn depends on the distribution of its components X_a and X_b. The support of $X_a = \frac{1}{2}\frac{N}{M}$ is $\left[0, \frac{1}{2}\right]$, the support of $X_b = \frac{1}{2}f$ is $\left[0, \frac{1}{2}\right]$, and the support of X is $[0, 1]$. For simplicity, we assume that X_a and X_b are independent with respective densities $p_a(x)$ and $p_b(x) = 2p_f(2x)$ where p_f is the density of $f \in [0, 1]$. The density of X is then given by the convolution

$$p(x) = \int_0^{x \wedge 1/2} p_a(x - t)\, p_b(t)\, dt \tag{8}$$

The density $p_a(x)$ of X_a is complicated by a point mass at $x = \frac{1}{2}$ arising from the upper bound of $N = \min(Y, M)$ where $Y = 2n_1 + n_2$. It is convenient to let (n_1, n_2) have a continuous joint density $p_{12}(a,b)$ over $[0,\infty)\times[0,\infty)$. This distribution can readily be transformed into a discrete distribution by rounding up or down noninteger values of (n_1,n_2) and obtaining the corresponding discrete probability distribution over $\mathbb{N}\times\mathbb{N}$ by integration over rectangles covering the integers. It is easier to work with the continuous version, and under these assumptions, the density of Y is

$$p_Y(y)=\frac{1}{2}\int_0^\infty p_{12}(0.5(y-b),b)\,db.$$

Since $N=\min(Y,M)=Y1_{\{Y<M\}}+M1_{\{Y\geq M\}}$, the distribution of N is mixed continuous-discrete with a rectified (spike and smooth) density

$$p_N(y)=p_Y(y)\times 1_{\{0\leq Y<M\}}+\mu_M\times\delta(y-M), \tag{9}$$

where $\mu_M=P\{Y\geq M\}=\int_M^\infty p_Y(y)\,dy$ and $\delta(x)$ is the Dirac delta function. The cdf of N is

$$P_N(y)=\int_0^{y\wedge M} p_Y(s)\,ds+\mu_M\times U(y-M),$$

where $U(x)$ is the step function $U(x)=1_{\{x\geq 0\}}$. The implied (rectified) density of $X_a=\frac{1}{2}\frac{N}{M}$ is $p_a(x)=2Mp_N(2Mx)$, which has the explicit form

$$p_a(x)=2Mp_Y(2Mx)\times 1_{\left\{0\leq x<\frac{1}{2}\right\}}+\mu_M\times\delta\left(2M\left(x-\frac{1}{2}\right)\right), \tag{10}$$

with a point mass of μ_M at $x=1/2$.

Combining (8) and (10), the density $p_X(x)$ over $x\in[0,1]$ is given by

$$\begin{aligned}
&\int_0^{x\wedge 1/2} p_a(x-t)\,p_b(t)\,dt\\
&=2M\int_0^{x\wedge 1/2} p_Y(2M(x-t))\,1_{\left\{0\leq x-t<\frac{1}{2}\right\}}p_b(t)\,dt\\
&\quad+\mu_M\times\int_0^x \delta\left(2M\left(x-t-\frac{1}{2}\right)\right)p_b(t)\,dt\\
&=2M\int_0^{x\wedge 1/2} p_Y(2M(x-t))\,p_b(t)\,dt+\mu_M p_b\left(x-\frac{1}{2}\right)\times 1_{\left\{x\geq\frac{1}{2}\right\}},
\end{aligned} \tag{11}$$

yielding (7). Observe that $p_Y(2M(x-t)) = 0$ for $x - t < 0$ and $p_b(t) = 0$ for $t > 1/2$, so that

$$\int_0^1 p_X(x)\,dx = 2M\int_0^1\int_0^{x\wedge 1/2} p_Y(2M(x-t))\,p_b(t)\,dtdx + \mu_M\int_{0.5}^1 p_b\left(x-\frac{1}{2}\right)dx$$

$$= 2M\int_0^{1/2}\int_0^x p_Y(2M(x-t))\,p_b(t)\,dtdx$$

$$+2M\int_{1/2}^1\int_0^{1/2} p_Y(2M(x-t))\,p_b(t)\,dtdx + \mu_M\int_0^{0.5} p_b(s)\,ds$$

$$= \int_0^{1/2}\int_0^{2M\left(\frac{1}{2}-t\right)} p_Y(s)\,ds p_b(t)\,dt + \int_0^{1/2}\int_{2M\left(\frac{1}{2}-t\right)}^M p_Y(s)\,ds p_b(t)\,dt + \mu_M$$

$$= \int_0^{1/2}\int_0^M p_Y(s)\,p_b(t)\,dtdx + \mu_M = \int_0^M p_Y(s)\,dx + \mu_M = 1 - \mu_M + \mu_M = 1.$$

A.2. Exact Theory

An exact theory suitable for computation can be obtained under explicit distributional assumptions concerning the primitive components (n_1, n_2, f) that determine the objective and subjective elements in X. We work with continuous distributions and simple parameterizations so that it is convenient to explore how different distributional shapes in the primitives impact the merit threshold distribution.

Let $n_1 \sim \Gamma(\kappa_1, \theta_1)$, $n_2 \sim \Gamma(\kappa_2, \theta_2)$, and $f = \frac{1}{\#(S_{all})}\sum_{j\in S_{all}} f_j \sim B(\alpha, \beta)$. Here $\Gamma(\kappa, \theta)$ denotes the gamma distribution with scale parameter $\theta > 0$, shape parameter $\kappa > 0$, and density $p(x) = \frac{1}{\Gamma(\kappa)\theta^\kappa}x^{\kappa-1}e^{-x/\theta}$ for $x \geq 0$, with mean $\mathbb{E}(n) = \kappa\theta$ and standard deviation $\mathbb{S}(n) = \kappa^{1/2}\theta$; and $B(\alpha, \beta)$ denotes the beta distribution with parameters $\alpha, \beta \geq 0$ and density $p(x) = \frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}$ for $x \in [0, 1]$, with mean $\mathbb{E}(f) = \alpha/(\alpha+\beta)$ and standard deviation $\mathbb{S}(f) = (\alpha\beta)^{1/2}/\left\{(\alpha+\beta)^2(\alpha+\beta+1)\right\}^{1/2}$.

The distribution of $Y = 2n_1 + n_2$ is the sum $\Gamma(\kappa_1, 2\theta_1) + \Gamma(\kappa_2, \theta_2)$. Upon convolution of these two gamma distributions and after some calculation, we obtain the density of Y

$$p_Y(x) = \frac{x^{\kappa-1}e^{-x/2\theta_1}}{\Gamma(\kappa)(2\theta_1)^{\kappa_1}\theta_2^{\kappa_2}}\sum_{j=0}^\infty \frac{(\kappa_2)_j}{j!(\kappa)_j}\left(\frac{1}{2\theta_1}-\frac{1}{\theta_2}\right)^j x^j$$

$$= \frac{e^{-x/2\theta_1}x^{\kappa-1}}{\Gamma(\kappa)(2\theta_1)^{\kappa_1}\theta_2^{\kappa_2}}\,{}_1F_1\left(\kappa_2, \kappa; x\left\{\frac{1}{2\theta_1}-\frac{1}{\theta_2}\right\}\right), \tag{12}$$

where ${}_1F_1$ is a confluent hypergeometric function, with $\kappa = \kappa_1 + \kappa_2$, and where we take $\theta_2 > 2\theta_1$ (a similar formula holds when $\theta_2 < 2\theta_1$). When $\theta_2 = 2\theta_1 = \theta$ the density is simply a gamma distribution with composite parameters (κ, θ).

The distribution of $X = X_a + X_b = \frac{1}{2}\frac{N}{M} + \frac{1}{2}f$, where $N = \min(Y, M)$, can now be obtained by quadrature using (7) upon specification of the parameters. The parameters can be classified as follows: (i) density parameters $\kappa_1, \theta_1, \kappa_2, \theta_2$, α, β that govern accomplishment and peer assessment; and (ii) control parameters M, γ_ℓ, γ_u that implement policy concerning winsorizing bibliometric data via the upper bound M and the upper γ_u and lower γ_ℓ limits to the overall assessment factor X which determine the merit thresholds.

A.3. Numerical Illustrations

The distribution (12) can be used for computation given explicit parameter values for the determining densities and the control parameters. We illustrate with the following classifications shown in Table 1 of the parameters corresponding to a selection of accomplishment and peer review levels ($\mathbb{E}$ and $\mathbb{S}$ denote mean and standard deviation of the respective distributions).

The high and low peer review parameters give mirror image densities for the peer review variate X_b on $[0, 0.5]$ and the mixed peer review parameters correspond to a uniform density, as shown in Fig. 2. The distribution of the bibliometric variate $Y = 2n_1 + n_2$ is calculated using (12), and the densities are shown in Fig. 3 for high, mixed, and low levels of accomplishment. The overall variate $X = X_a + X_b$ has density $p_X(x)$ which is computed using (11). The densities are shown in Fig. 1 (given earlier in the paper) for high accomplishment (HA) combined with high peer review (HR), mixed peer review (MR), and low peer review (LR). Figs. 4–5 show the corresponding densities for a mixed level of accomplishment (MA) and low accomplishment (LA). The merit threshold distribution of τ is computed using the rectified (double spike and smooth) density $p_\tau(t)$ given in (5) and (6). Table 2 presents summary statistics calculated for this merit threshold distribution, showing the probability $P(\tau \geq \tau_U)$ of exceeding the upper threshold τ_U, the probability $P(\tau \leq \tau_L)$ of exceeding the lower threshold τ_L, and the mean threshold level $\mathbb{E}(\tau)$.

TABLE 1
Parameter Classifications

Peer Review	*High*	*Mixed*	*Low*
(α, β)	$\alpha = 5, \beta = 1$	$\alpha = 1, \beta = 1$	$\alpha = 1, \beta = 5$
$\mathbb{E}(f)$ $\mathbb{S}(f)$	0.83 0.14	0.5 0.28	0.16 0.14
Accomplishment	*High*	*Mixed*	*Low*
$(\kappa_1, \kappa_2, \theta_1, \theta_2)$	$(2, 8, 2, 4)$	$(2, 2, 2, 3)$	$(1, 2, 0.5, 2)$
$\mathbb{E}(n_1)$ $\mathbb{E}(n_2)$	8 32	4 6	0.5 4
$\mathbb{S}(n_1)$ $\mathbb{S}(n_2)$	2.83 11.31	2.82 4.24	0.5 2.82
Controls	$(M, \gamma_1, \gamma_2) = \left(50, \frac{1}{4}, \frac{3}{4}\right)$		

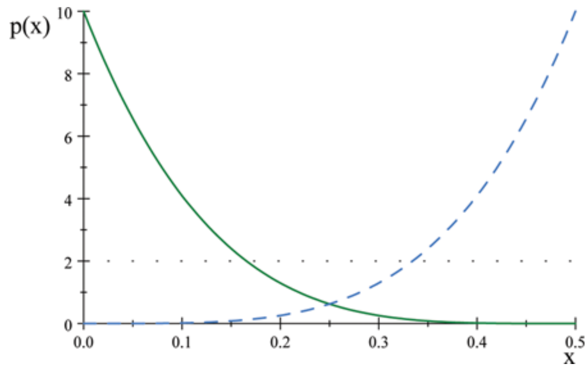

FIGURE 2 Peer review (X_b) density $p_b(x) = \frac{2(2x)^{\alpha-1}(1-2x)^{\beta-1}}{B(\alpha,\beta)}$ for low peer review $\{\alpha = 1, \beta = 5\}$ (solid/green), mixed peer review $\{\alpha = 1, \beta = 1\}$ (dotted/black), and high peer review $\{\alpha = 5, \beta = 1\}$ (dashed/blue)

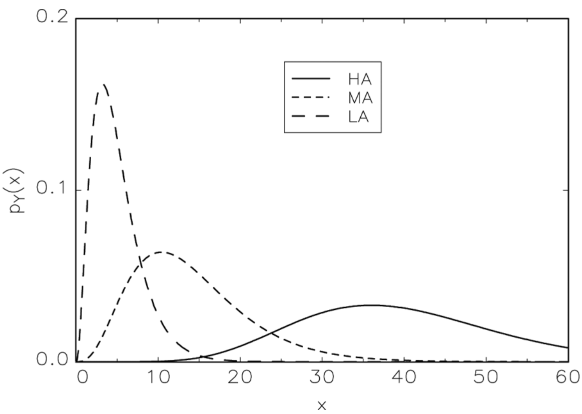

FIGURE 3 Densities of the bibliometric component $Y = 2n_1 + n_2$ for parameter values corresponding to high (HA), mixed (MA), and low (LA) levels of accomplishment given in Table 1

High accomplishment and high peer review produce a density for overall accomplishment X that is concentrated in the upper part of the interval $[0, 1]$, which leads to a high probability $P(\tau \leq \tau_L) = 0.762$ of reaching the lower threshold τ_L and makes fellowship election easier. The mean threshold level in this case is $\mathbb{E}(\tau) = 0.236$, close to the lower bound control parameter $\tau_L = 0.2$. Fig. 1 shows discontinuities in the density $p_X(x)$ in two cases (HA and MR; HA and LR) which arise from a nonnegligible probability $P(X_a \geq M)$ of X_a exceeding the control parameter bound M which delimits quantity effects in bibliometric data to that level. In each of these cases the density $p_b\left(x - \frac{1}{2}\right) > 0$ at $x = \frac{1}{2}$, thereby producing the discontinuity in $p_X(x)$. In the high peer review case (HA and HR), $p_b\left(x - \frac{1}{2}\right) = 0$ at $x = \frac{1}{2}$ and the density $p_X(x)$ is continuous.

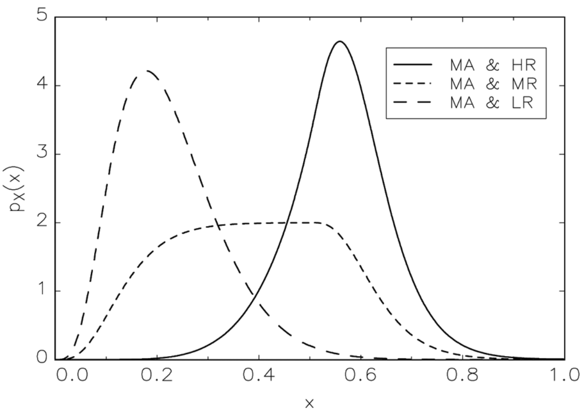

FIGURE 4 Densities of $X = X_a + X_b$ for mixed accomplishment (MA) and high (HR), mixed (MR), and low (LR) peer review

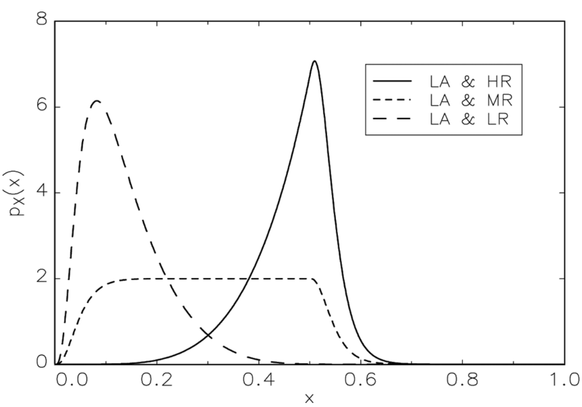

FIGURE 5 Densities of $X = X_a + X_b$ for low accomplishment (LA) and high (HR), mixed (MR), and low (LR) peer review.

In a similar way, low accomplishment and low peer review produce a density for X that is concentrated in the lower part of the interval $[0, 1]$, giving a high probability $P(\tau \geq \tau_U) = 0.912$ of exceeding the upper threshold τ_U and a zero probability of reaching the lower threshold τ_L, making fellowship election harder because of the high voting threshold required for election. In this case, the mean threshold level is $\mathbb{E}(\tau) = 0.496$, which is very close to the upper bound control parameter $\tau_U = 0.5$. Table 2 provides a selection of other cases, showing how mixtures of high and low levels of accomplishment and peer reviews affect the merit threshold.

TABLE 2
Merit Threshold Statistics

Peer Review	*High*			*Mixed*			*Low*		
Accomplishment	*High*	*Mixed*	*Low*	*High*	*Mixed*	*Low*	*High*	*Mixed*	*Low*
$P(\tau \geq \tau_U)$	0.000	0.003	0.013	0.008	0.227	0.337	0.031	0.651	0.912
$P(\tau \leq \tau_L)$	0.762	0.028	0.000	0.371	0.008	0.000	0.050	0.000	0.000
$\mathbb{E}(\tau)$	0.236	0.316	0.369	0.323	0.406	0.445	0.421	0.482	0.496

A.4. Fellowship Elections in Economics, Statistics and National Academies

This Appendix provides some background information on fellowship election or appointment procedures as they are currently performed in various leading societies in economics, statistics and the natural sciences.

Econometric Society: Annual fellowship elections are held and the electorate comprises existing fellows of the society. (Prior to 1960 fellows were nominated and elected by the council.) Names of candidates nominated for election are placed on a ballot and fellows return a Yes/No vote. To secure election, candidates must obtain 30% or more votes in the election. Nominations are by petition of at least three members of the society (who are usually, but not necessarily, fellows) or by a nominating committee appointed by the president. Nominations include a statement of the candidate's contributions, a list of up to six major publications, reference to the candidate's home webpage, and a list of those nominating the candidate and an indication whether the nominating committee endorses the candidate. "To be eligible for nomination as a Fellow, a person must have published original contributions to economic theory or to such statistical, mathematical, or accounting analyses as have a definite bearing on problems in economic theory, and must be, or upon election become, a member of the Society."[23]

American Economic Association: Distinguished fellowships are by special appointment. "Past Presidents of the Association shall be Distinguished Fellows. Additional Distinguished Fellows may be elected, but not more than three in any one calendar year from economists of high distinction in the United States and Canada."[24]

European Economic Association: Fellows are elected by virtue of the office held in the Association. Fellowships are "bestowed on the Association's officers, the editors of the Association's journal, the Programme Chairs of its annual Congresses, as well as the Marshall and Schumpeter lecturers. Becoming a Fellow is contingent on becoming a member of the association."[25]

[23]Econometric Society website is http://www.econometricsociety.org/society.asp.
[24]American Economic Association website is http://www.vanderbilt.edu/AEA/disting_fellows.htm.
[25]European Economic Association website is http://www.eeassoc.org/index.php?page=21.

Institute of Mathematical Statistics: Election is by a special fellows selection committee which reviews nominations. Qualification for fellowship requires "demonstrated distinction in research in statistics or probability by publication of independent work of merit" or "well-established leadership whose contributions to the field of statistics or probability ... or the application of statistics or probability ... shall be judged of equal value."[26]

Royal Statistical Society: No merit assessment is required. "Fellowship is open to all who have an interest in statistics: formal qualifications are not needed."[27]

American Statistical Association: Election is by a fellows selection committee which reviews nominations that require online submission of detailed forms about the candidate, letters of support (at most four), draft citations, and other information. "Each committee member assigns a rating from 1 to 5 to a given nominee, with 1 being the lowest and 5 being the highest; non-integer ratings are perfectly acceptable. Though there are no fixed criteria for rating a nomination, the following table provides some examples of how a rater might typically react to a nomination package."[28] This table[29] indicates the potential impact of various criteria on a committee member's subjective assessment of a nominated candidate. For example, "sole authorship of 5 or more articles in leading statistical journals", "strong evidence of positive impact of mentoring," and "Program committee chair for a major ASA meeting"[30] are all rated as "++" in terms of impact.

The Royal Society: There is an upper limit of 44 new Fellows, 8 Foreign Members and 1 Honorary Fellow. Candidates for the Fellowship or Foreign Membership must be nominated by two Fellows of the *Royal Society*, who sign a certificate of proposal. "The Council of the *Royal Society* oversees the selection process. Two Officers, the Biological Sciences Secretary and the Physical Sciences Secretary, are responsible for the smooth running of this process. The Council appoints ten subject area committees, known as Sectional Committees, to advise it about the selection of the list of the strongest candidates. Each candidate is considered by the relevant Sectional Committee on the basis of a full curriculum vitae, details of their research achievements, a list of all their scientific publications and a copy of their 20 best scientific papers. Members of the Sectional Committees vote to produce a short-list. The final list of candidates is confirmed by the Council and a secret ballot of Fellows is held. A candidate is elected if he or she secures two-thirds of votes."[31]

The British Academy: Fellowship is by election and Academy Council decision. "Candidates are proposed by Section Standing Committees, Fellows or by Vice Chancellors and Principals of U.K. Universities. Section Committees meet to agree the

[26]IMS official website is http://www.imstat.org/awards/fellows.htm.
[27]Royal Statistical Society website is http://membership.rss.org.uk/main.asp?page=1280.
[28]American Statistical Association website is http://www.amstat.org/.
[29]See http://www.amstat.org/fellows/nominations/pdfs/RatingofNominees.pdf.
[30]American Statistical Association website is http://www.amstat.org/careers/fellows.cfm.
[31]The Royal Society website is http://royalsociety.org/about-us/fellowship/election/.

names to be put to a secret ballot within each Section. Sections agree the names of assessors. The case for election and the particular distinction of each candidate is sent to assessors. The reports of assessors supply an independent judgment to supplement and inform the deliberations of the Section. A secret ballot is then conducted within each Section. Sections meet to study the ballot results and to make recommendations to the humanities and social science Groups, which are responsible for ensuring consistency across Sections. The Groups' and the Fellowship and Structures Committee's recommendations are considered and discussed by the Academy's Council, which agrees a list of names to be nominated for election to the Annual General Meeting of Fellows."[32]

The National Academy of Sciences: Membership is by election. "Only Academy members may submit formal nominations. Consideration of a candidate begins with his or her nomination, followed by an extensive and careful vetting process that results in a final ballot at the Academy's annual meeting. Currently, a maximum of 84 members may be elected annually."[33]

ACKNOWLEDGMENT

My thanks to four referees, the Editor Essie Maasoumi, Steven Durlauf, John Rust, and Peter Schmidt for comments on earlier drafts and some useful additional references. Olav Bjerkholt kindly guided me to original source material related to Ragnar Frisch and the Econometric Society.

FUNDING

Support from a Kelly Fellowship and the NSF under Grant No. SES 09-56687 is gratefully acknowledged.

REFERENCES

Acemoglu, D., Robinson, J. (2000). Why did the west extend the franchise? Democracy, inequality and growth in historical perspective. *Quarterly Journal of Economics* 115:1167–1199.

Andvig, J. C., Thonstad, T. (1998). Ragnar Frisch at the University of Oslo. Chapter 1. In Strom, S., ed. *Econometrics and Economic Theory in the 20th Century. The Ragnar Frisch Centennial Symposium*. Cambridge: Cambridge University Press, pp. 3–25.

Atkinson, A. B. (2003). Multidimensional deprivation: Contrasting social welfare and counting approaches. *Journal of Economic Inequality* 1:51–65.

Atkinson, A. B., Bourguignon, F. (1982). The comparison of multi-dimensioned distributions of economic status. *Review of Economic Studies* 49:183–201.

Baltagi, B. H. (2007). Worldwide Econometrics Rankings: 1989 2005. *Econometric Theory* 23:952–1012.

[32] British Academy website is http://www.britac.ac.uk/fellowship/elections/elecproc.cfm.

[33] National Academy of Sciences website is http://www.nasonline.org/.

Bjerkholt, O. (1995). Ragnar Frisch, the Originator of Econometrics, Introduction. In: Bjerkholt, O., ed. *Foundations of Modern Econometrics: The Selected Essays of Ragnar Frisch*. Vol. 1. Edward Elgar, pp. xiii-lii.

Bjerkholt, O. (1998). Ragnar Frisch and the foundation of the econometric society and *Econometrica*. Chapter 2. In Strom, S., ed. *Econometrics and Economic Theory in the 20th Century. The Ragnar Frisch Centennial Symposium*. Cambridge University Press, pp. 26–57.

Bjerkholt, O. (2005). Frisch's econometric laboratory and the rise of Trygve Haavelmo's probability approach. *Econometric Theory* 21:491–533.

Bjerkholt, O., Qin, D. (2010). Editors' introduction. Teaching economics as a science: The 1930 Yale lectures of Ragnar Frisch. In Bjerkholt, O., Qin, D., eds. *A Dynamic Approach to Economic Theory: The Yale Lectures of Ragnar Frisch*, 1930. London: Routledge, pp. 1–28.

Bjerkholt, O. (2015). Trygve Haavelmo at the Cowles Commission., *Econo-metric Theory*, 2015 (forthcoming).

Brooks, T. (1986). Evidence of complex citer motivations. *Journal of the American Society for Information Science* 37:34–36.

Chang, C.-L., McAleer, M., Oxley, L. (2011). Great expectatrics: great papers, great journals, great econometrics. *Econometric Reviews* 30:583–619.

Cozzens, S. E. (1989). What do citations count? The rhetoric-first model. *Scientometrics* 15:437–447.

Gowdy, J. (2006). Hunter gatherers and the mythology of the market. In: Lee, R. B., Daly, R. H., eds. *The Cambridge Encyclopedia of Hunters and Gatherers*. Cambridge: Cambridge University Press.

Hamermesh, D. S., Schmidt, P. (2003). The determinants of econometric society fellows elections. *Econometrica* 71:399–407.

Hall, P., Miller, H. (2010). Modeling the variability of rankings. *Annals of Statistics* 38:2652–2677.

Hall, P. (2011). 'Ranking our excellence' or 'assessing our quality' or whatever .... *Institute of Mathematical Statistics Bulletin* 40:12–14.

Louça, F. (2007). *The Years of High Econometrics*. London: Routledge.

Louça, F., Terlica, S. (2011). The fellowship of econometrics. Selection and diverging views in the province of mathematical economics from the 1930s to the 1950s. *History of Political Economy*, Vol. 43. (Annual Supplement), pp. 57–85.

Luganoff, R. (2009). Dynamic stability and reform of political institutions. *Games and Economic Behavior* 67:569–583.

Maasoumi, E. (1999). Multidimensioned approaches to welfare analysis. In: Silber, J., ed. *Handbook of Income Inequality Measurement*. New York: Kluwer Publishers, pp. 437–484.

Phillips, P. C. B. (2005). Automated discovery in econometrics. *Econometric Theory* 21:3–20.

Roberts, K. (1999). Dynamic voting in clubs. Suntory Toyota Centre Working paper # 367. London School of Economics.

Schultz, H. (1924). The statistical measurement of the elasticity of demand for beef. *Journal of Farm Economics* 6:254–78.