MERITOCRACY VOTING: MEASURING THE UNMEASURABLE

By

Peter C.B. Phillips

October 2011

COWLES FOUNDATION DISCUSSION PAPER NO. 1833

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

http://cowles.econ.yale.edu/

Meritocracy Voting: Measuring the Unmeasurable

Peter C. B. Phillips*

Yale University, University of Auckland,

University of Southampton & Singapore Management University

October 2011

Abstract

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 judgements 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 judgement 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 paper 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 judgements 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. The methodology has application to a wide arena of quality assessment and professional ranking exercises.

Keywords: Bibliometric data, Election, Fellowship, Measurement, Meritocracy, Peer review, Quantification, Subjective assessment, Voting

JEL classification: A14, Z13, C18

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

"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 Aristotle: 'Measure the Unmeasurable' ", Ragnar Frisch (Examination report as a student at the University of Oslo, cited in Louça, 2007.)

1 Introduction

Hierarchical elements and status inequalities are pervasive in modern industrialized society. Social stratifications arise from multiple sources such as socioeconomic 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 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 accordingly is highly prized.

The subject of the present paper is the selection process by which such fellowships are determined. Assessment of merit necessarily involves human judgement about the contributions of individual candidates. But information about and opinions of those contributions may differ considerably in a voting population. Debate on the qualifications for fellowship are as ancient as learned societies themselves. In an archival study on the foundation of the Econometric Society, for example, Louça and Terlica (2011) provide extensive evidence of diverging views among the founders of that Society about candidates for fellows in the early 1930s. They report continuing divisive debates among the broader fellowship in the 1950s about selection criteria for fellowship[1].

Distortions in voting may arise for many reasons. For instance, intellectual founders and leaders may veto certain candidates[2]; 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? What elements - quantitative and qualitative - might enter into the selection process to substantiate 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?

Hamermesh and Schmidt (2003) analyzed data from fellowship elections in the Econometric Society over 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.

Finding a mechanism for promoting fairness across fields, institutions and regions, collecting and distributing the relevant information that can assist in

[1]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 PC Mahalanobis) and opposed by others (including Tjalling Koopmans and Jacob Marschak). In the end, no changes to criteria or procedures for fellowships were made.

[2]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" (op.cit. p. 31), an injustice as it turned out. Another candidate was repeatedly opposed as president of the society as "not recommendable" (op.cit. p. 35) on grounds that he "uses many words to express his meanings" (op.cit. p. 35). See Louça and Terlica (2011) for further examples and discussion.

this process, and respecting subjective assessments of credentials across a population of electors is a serious challenge 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.[3]

This paper seeks to offer some material assistance toward that goal. It provides a quantitative rule that combines human judgement 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 UK, Australia and New Zealand) are one example. Journal rankings and impact factors of research are another. 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 data and search engine methods. The challenge we face in such assessment exercises is to utilize the vast and growing quantity of bibliometric 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 paper 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

[3]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.

fellows in the society, a fellowship selection committee or even the governing body or council. Some examples from leading learned societies in economics, statistics and national academies are given in the Appendix. The threshold 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 chair.

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. 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 paper 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 judgement. 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 judgemental – 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 judgements 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.

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. For certain explicit distributions of objective and subjective evidence, we will report some exact distributional results that show the response distribution of τ to its inputs. These distributions reveal the flexibility of the approach and the way different types and levels of credential information contribute to outcomes.

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 judgemental views of the accomplishment. What follows is one possible 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. In the following section 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 component X_a depends on quantitative information about research accomplishment and material contribution to the discipline. For example, 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 what 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. 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.

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. For the formula given here, we use the following approach. Each member (j) of the voting electorate reports a 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). These assessments are averaged to produce a subjective accomplishment factor $f = \frac{1}{\#(S_{all})} \sum_{j \in S_{all}} f_j$ 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:

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

where

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

and S_{valid} is the set of electors who returned a valid assessment factor $\varphi \in (0, 1)$. According to this rule, abstentions, non voters 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

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

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 forces 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 assessments may be controlled in X_b by adjusting the support of φ.

Having determined each voter's f_j by this process, the aggregate component $f = \frac{1}{\#(S_{all})} \sum_{j \in S_{all}} f_j$ represents the average subjective view of the voting electorate on the candidate. The subjective contribution to the accomplishment factor X is then $X_b = \frac{1}{2} \times f$, whose support is $U_b = \left[0, \frac{1}{2}\right]$, and X_b carries an implicit weight of $\frac{1}{2}$ in X. Importantly, X_b places demands on individual electors that go beyond simple Yes/No or rank voting schemes. Each elector must translate a subjective judgement of a candidate into a quantitative subjective score φ_j for that candidate. If this score is to count then the elector must choose a $\varphi_j \in (0,1)$. Otherwise a community based score φ will be used instead[4]. 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 and by a direct Yes/No or rank order vote on the candidate.

Based on these two components the overall accomplishment factor is computed as $X = X_a + X_b$. 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

$$\tau = \tau_U 1_{\{X<\gamma_\ell\}} + \tau_L 1_{\{X>\gamma_u\}} + \left[\tau_L + (\tau_U - \tau_L)\left\{\frac{\gamma_u - X}{\gamma_u - \gamma_\ell}\right\}\right] 1_{\{\gamma_\ell \leq X \leq \gamma_u\}}, \tag{3}$$

where 1_A is the indicator of A. For each candidate (i) in the election the corresponding merit threshold τ_i is computed using formula (3) 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 (A_i) supporting that candidate in the election. If $A_i \geq \tau_i$ so that the percentage of actual votes meets or exceeds the candidate's merit threshold (τ_i) then the

[4]Other possibilities might usefully be considered. For example, field differentials may mean that some electors have difficulty appreciating (or even reading) a candidate's work when it is far from their own field of research. In such cases, an elector may not be comfortable returning a personal peer review of the candidate, but may be ready to delegate the assessment to others, such as (i) the nominating group, (ii) the fellows selection committee (if one exists), or (iii) all other fellows. Assignation to these alternatives may be arranged by an elector being offered distinct discrete choices that signal these assigned assessments, such as (i) $\varphi_j = 2$, (ii) $\varphi_j = 3$, and (iii) $\varphi_j = 1$ or 0, as in rule (1).

candidate is elected. Symbolically, $E_i = 1_{\{A_i \geq \tau_i\}}$ gives the election outcome (1 = success; 0 = failure).

In practice, the main effect of (3) is to require a higher percentage of votes in the election for candidates with less demonstrated accomplishment, as represented 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 7.3 of the Appendix).

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 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:

(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 $f = \frac{1}{\#(S_{all})} \sum_{j \in S_{all}} f_j \in [0, 1]$ with f_j determined as in (1).

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. For example, 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 (3) however, the peer view of the entire electorate is taken into account in the measurement of X_b 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 tends to dilute the impact of coalition voting in the election.

4 The Merit Threshold Distribution

The 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, 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

$$p_U = P(\tau \geq \tau_U) = \int_0^{\gamma_\ell} p_X(x)\, dx, \quad p_L = P(\tau \leq \tau_L) = \int_{\gamma_u}^1 p_X(x)\, dx, \qquad (4)$$

and the density by

$$p_\tau(t) = p_X\left(\gamma_\ell + \frac{\gamma_u - \gamma_\ell}{\tau_U - \tau_L}(\tau_U - t)\right)\frac{\gamma_u - \gamma_\ell}{\tau_U - \tau_L}, \quad \text{for } t \in (\tau_L, \tau_U). \qquad (5)$$

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 X_a and X_b are independent, then the distribution of X has the convolution form

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

With this structure it is possible to obtain the density $p_X(x)$ in terms of the density[5] $p_Y(y)$ of Y and the density $p_b(x)$ of X_b. Derivations are given in section 7.1 of the Appendix, where it is shown that

$$p_X(x) = 2M \int_0^{x \wedge 1/2} p_Y(2M(x-t))\, p_b(t)\, dt + \mu_M \times p_b\left(x - \frac{1}{2}\right) \times 1_{\{x \geq \frac{1}{2}\}}. \tag{6}$$

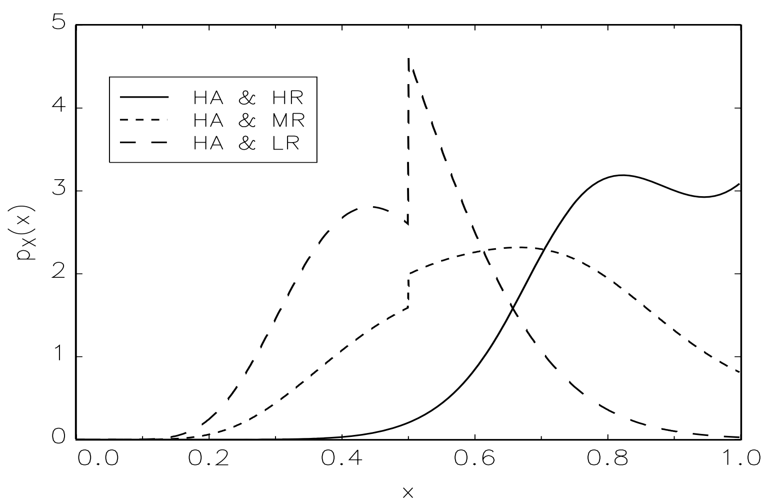

Fig. 1: Densities of $X = X_a + X_b$ for distributions of X_a and X_b corresponding to high accomplishment (HA) and high peer review (HR), mixed peer review (MR), and low peer review (LR).

The density $p_X(x)$ can have a (discrete) 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

[5] 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.

that an individual's 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 paper (Section 7). 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 & LR) case in the figure.

5 Implementation

Formula (3) 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 judgements that underlie publication. The 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 can be coordinated by a society's governing body and is readily accomplished online using a web server. The key steps are detailed below.

1. **Prior Parameter Settings.** Parameters that appear in formula (3) 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 (3).

   (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 (3)) 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 (3)) which allocates a weight of λ to quantitative information X_a and a weight of $1 - \lambda$ to subjective assessment X_b.

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.

   (a) A nominating statement of some designated length (such as 200 words).

   (b) A list of n^* of the candidate's most influential publications (typically, $n^* \leq 5$).

   (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. Deadline for the submission of all nominations, including the objective and subjective information that must accompany the nomination.

4. A criterion 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 separate nominations (typically $n^{\#} \geq 3$).

5. Deadline for the online distribution to the voters of information about all nominated candidates for election. This information will include the nomination information $2a, 2b, 2c$ listed above, together with the objective

component X_a and the subjective component $X_b^{\text{nom}} = \frac{1}{2} \times f^{\text{nom}}$, where $f^{\text{nom}} = \frac{1}{\#(S_{nom})} \sum_{k \in S_{nom}} \varphi_k$ 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.

6. Deadline for the election votes and assessments to be submitted. These votes include both the vote itself (Yes/No) *and* the subjective assessment factor made by each elector for each nominated candidate.

7. After the election, votes for each candidate are counted. The subjective assessment data submitted in the election returns by each voter is 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 (A_i) of all valid votes cast and compared with the candidate's merit threshold τ_i. Candidate i is elected if $A_i \geq \tau_i$.

Some computations that are given later in this paper 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 (3), elections require some parameter settings. 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 then any information in X is ignored.

The quantitative information (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.

6 Conclusion

The focus of the present contribution is the appraisal of credentials, the operational use of available quantitative information and the pooling of human judgement 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 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 dubbed 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 nothing more than a first step in addressing this issue. The specific rule (3) 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 provide information about received opinion on research (or its neglect). In both cases, 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. As we have discussed, voters end up influencing the election of each candidate in two separate ways. They report their subjective assessment of a candidate into a numerical score that combines with the judgements of other voters and material information about the candidate to determine the candidate's merit threshold. They 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. Both the subjective assessment and the vote influence the final outcome.

The formulation given here is a tentative beginning. Obviously a great deal more work can go into its further mathematical development, into the use of voting theory in its formulation, and into its online implementation. More attention to data sources and the quantification of subjective assessment both 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 paper will have achieved its limited goal if it serves to stimulate further thinking on these issues and on the general problem of quantifying the assessment of merit. Research on this topic seems important not only for learned society decision making but for the many other instances in academic life where merit assessment is such a critical matter in the careers and lives of our colleagues.

7 Appendix

7.1 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 $[0, 1]$. 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

$$p(x) = \int_0^{x \wedge 1/2} p_a(x - t)\, p_b(t)\, dt \tag{7}$$

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 non integer 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

$$p_N(y) = p_Y(y) \times 1_{\{0 \leq Y < M\}} + \mu_M \times \delta(y - M), \tag{8}$$

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

$$p_a(x) = 2Mp_Y(2Mx) \times 1_{\left\{0 \leq x < \frac{1}{2}\right\}} + \mu_M \times \delta\left(2M\left(x - \frac{1}{2}\right)\right), \tag{9}$$

with a point mass of μ_M at $x = 1/2$.

Combining (7) and (9), the density $p_X(x)$ over $x \in [0,1]$ is given by

$$\begin{aligned} & \int_0^{x \wedge 1/2} p_a(x-t)\, p_b(t)\, dt \\ = \quad & 2M \int_0^{x \wedge 1/2} p_Y(2M(x-t))\, 1_{\left\{0 \leq x-t < \frac{1}{2}\right\}} p_b(t)\, dt \\ & + \mu_M \times \int_0^x \delta\left(2M\left(x - t - \frac{1}{2}\right)\right) p_b(t)\, dt \\ = \quad & 2M \int_0^{x \wedge 1/2} p_Y(2M(x-t))\, p_b(t)\, dt + \mu_M \times p_b\left(x - \frac{1}{2}\right) \times 1_{\left\{x \geq \frac{1}{2}\right\}}, \end{aligned} \tag{10}$$

yielding (6). Observe that $p_Y(2M(x-t)) = 0$ for $x - t < 0$ and $p_b(t) = 0$ for $t > 1/2$, so that

$$\begin{aligned} & \int_0^1 p_X(x)\, dx = 2M \int_0^1 \int_0^{x \wedge 1/2} p_Y(2M(x-t))\, p_b(t)\, dt dx + \mu_M \int_{0.5}^1 p_b\left(x - \frac{1}{2}\right) dx \\ = \quad & 2M \int_0^{1/2} \int_0^x p_Y(2M(x-t))\, p_b(t)\, dt dx \\ & + 2M \int_{1/2}^1 \int_0^{1/2} p_Y(2M(x-t))\, p_b(t)\, dt dx + \mu_M \int_0^{0.5} p_b(s)\, ds \\ = \quad & \int_0^{1/2} \int_0^{2M\left(\frac{1}{2} - t\right)} p_Y(s)\, ds p_b(t)\, dt + \int_0^{1/2} \int_{2M\left(\frac{1}{2} - t\right)}^M p_Y(s)\, ds p_b(t)\, dt + \mu_M \\ = \quad & \int_0^{1/2} \int_0^M p_Y(s)\, p_b(t)\, dt dx + \mu_M = \int_0^M p_Y(s)\, dx + \mu_M = 1 - \mu_M + \mu_M = 1. \end{aligned}$$

7.2 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 following density of Y

$$\begin{aligned} p_Y(x) &= \frac{x^{\kappa-1} e^{-x/2\theta_1}}{\Gamma(\kappa)(2\theta_1)^{\kappa_1} \theta_2^{\kappa_2}} \sum_{j=0}^{\infty} \frac{(\kappa_2)_j}{j!(\kappa)_j} \left(\frac{1}{2\theta_1} - \frac{1}{\theta_2}\right)^j x^j \\ &= \frac{e^{-x/2\theta_1} x^{\kappa-1}}{\Gamma(\kappa)(2\theta_1)^{\kappa_1} \theta_2^{\kappa_2}} \, {}_1F_1\left(\kappa_2, \kappa; x\left\{\frac{1}{2\theta_1} - \frac{1}{\theta_2}\right\}\right), \end{aligned} \tag{11}$$

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 (6) upon specification of the parameters. The parameters can be classified as follows: (i) density parameters $\kappa_1, \theta_1, \kappa_2, \theta_2, \alpha, \beta$ 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.

7.3 Illustration

The distribution (11) 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.

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	$(M, \gamma_1, \gamma_2) = \left(50, \frac{1}{4}, \frac{3}{4}\right)$		

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 (11) 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 (10). 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 (4) and (5). 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 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

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 & MR; HA & 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 & HR), $p_b\left(x - \frac{1}{2}\right) = 0$ at $x = \frac{1}{2}$ and the density $p_X(x)$ is continuous.

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.

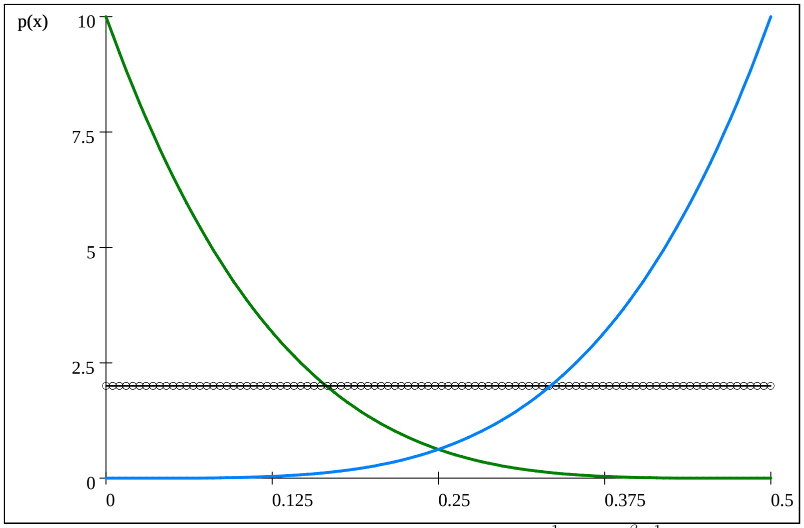

Fig. 2: Peer review (X_b) density $p_b\left(x\right) = \frac{2(2x)^{\alpha-1}(1-2x)^{\beta-1}}{B(\alpha,\beta)}$ for high peer review $\{\alpha = 1, \beta = 5\}$ (solid/green), mixed peer review $\{\alpha = 1, \beta = 1\}$ (dotted/black), and low peer review $\{\alpha = 5, \beta = 1\}$ (dashed/blue).

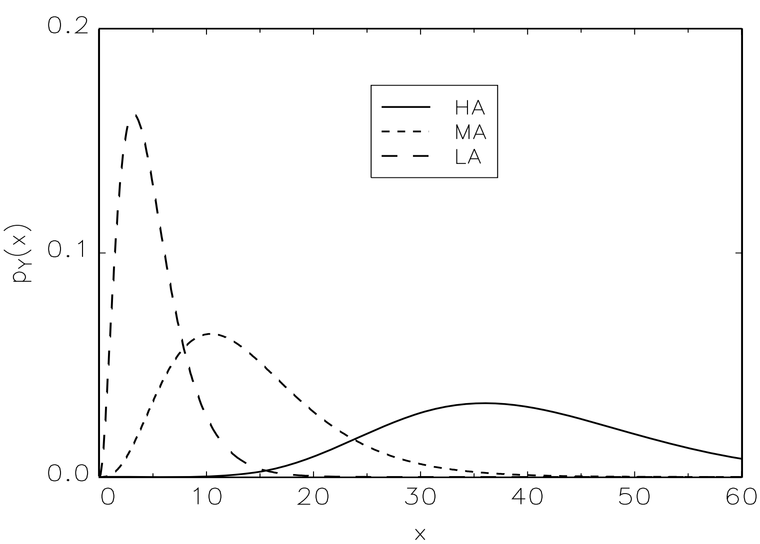

Fig. 3: Densities of the bibliometric component $Y = 2n_1 + n_2$ for parameter values corresponding to high (HA), mixed (MA) and low (LA) levels of accomplishment given in Table 1.

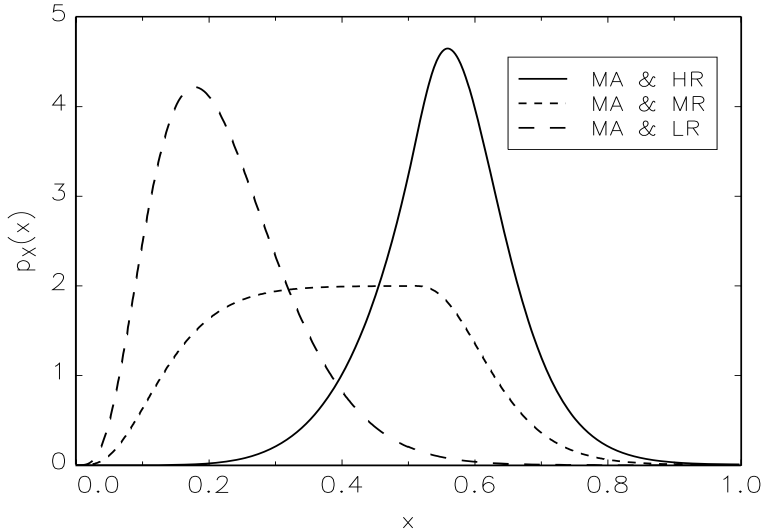

Fig. 4: Densities of $X = X_a + X_b$ for mixed accomplishment (MA) and high (HR), mixed (MR), and low (LR) peer review.

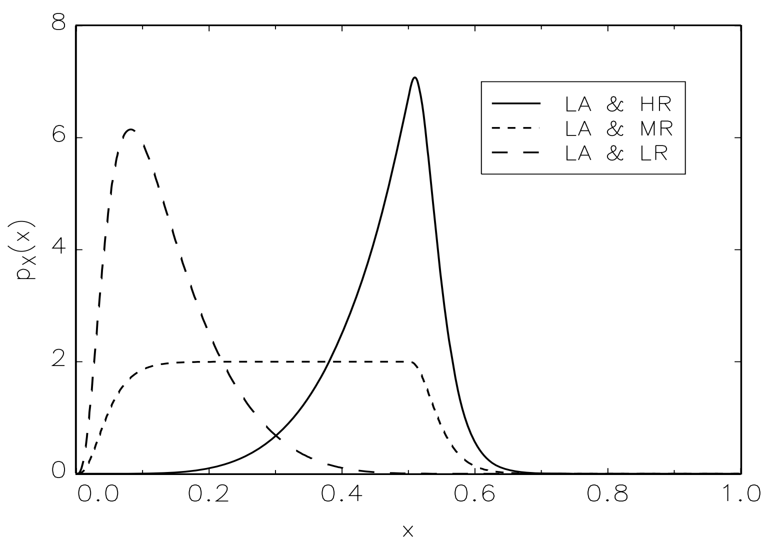

Fig. 5: Densities of $X = X_a + X_b$ for low accomplishment (LA) and high (HR), mixed (MR), and low (LR) peer review.

7.4 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."[6]

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."[7]

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."[8]

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"[9].

Royal Statistical Society: No merit assessment is required. "Fellowship is open to all who have an interest in statistics: formal qualifications are not needed."[10]

American Statistical Association: Election is by a fellows selection commit-

[6] Econometric Society website: http://www.econometricsociety.org/society.asp
[7] American Economic Association website: http://www.vanderbilt.edu/AEA/disting_fellows.htm
[8] European Economic Association Website: http://www.eeassoc.org/index.php?page=21
[9] IMS Official Website: http://www.imstat.org/awards/fellows.htm
[10] Royal Statistical Society website: http://membership.rss.org.uk/main.asp?page=1280

tee 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."[11] This table[12] 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"[13] 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."[14]

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 UK Universities. Section Committees meet to agree the 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 judgement 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."[15]

The National Academy of Sciences: Membership is by election. "Only Academy members may submit formal nominations. Consideration of a candidate

[11] American Statistical Association website: http://www.amstat.org/
[12] See http://www.amstat.org/fellows/nominations/pdfs/RatingofNominees.pdf
[13] American Statistical Association website: http://www.amstat.org/careers/fellows.cfm
[14] The Royal Society website: http://royalsociety.org/about-us/fellowship/election/
[15] British Academy website: http://www.britac.ac.uk/fellowship/elections/elecproc.cfm

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."[16]

8 References

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

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

Hamermesh, D. S and P. Schmidt 2003. "The determinants of econometric society fellows elections", *Econometrica*, 71, 399-407.

Louça, F. 2007. *The Years of High Econometrics*. Routledge.

Louça, F. and S. Terlica. 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* (forthcoming).

Luganoff, R. 2009. Dynamic stability and reform of political institutions. *Games and Economic Behavior*, 67, 569-583.

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

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