B A D A N I A   O P E R A C Y J N E   I   D E C Y Z J E

Nr 3–4 2007

Jacek W. MERCIK*

ECONOMETRIC ESTIMATION OF HIDDEN FACTORS IN GROUP DECISION MAKING – THEIR IMPACT ON POWER INDEX ESTIMATION[1]

Assessing the voting power of a decision maker depends on the appropriate evaluation of all factors influencing such power. Models constructed with the use of a set of given factors are usually statistically significant, but at the same time the determination coefficient is very low. An econometric technique is proposed, which increases the determination coefficient of such a model and consequently improves the estimation of voting power using a power index. Such an approach enables us to investigate the nature of hidden factors by considering their effects and variation in the model. An example and algorithm for improving the econometric model are also presented.

Keywords: *hidden factors, power index*

Assumptions used in the research: an example of econometric assessment

In the book of Gladysz and Mercik [5], one may find an example of modeling the voting frequency in the Polish presidential election of 2000. The econometric model obtained is:

$$\textit{Frequency} = 0.6597 - 0.1349\ \textit{unemployment rate} + 0.4832\ \textit{number of firms registered}$$

* Institute of Organization and Management, Wrocław University of Technology, Smoluchowskiego 25, 50-372 Wrocław, e-mail: Jacek.Mercik@pwr.wroc.pl

[1] An earlier version of this paper was presented during the SIND2 and XVI IMGTA Meeting in Foggia, Italy, June 14–17, 2006, during the Special Session "VOTING AND POWER INDICES", organized by Gianfranco Gambarelli and supported by the Department of Mathematics, Statistics, Computer Science and Applications "Lorenzo Mascheroni", University of Bergamo.

It is statistically significant, but has a relatively low determination coefficient: $R^2 = 0.3009; p - value = 8.98E - 20$. Therefore, there is a hidden factor (or factors) which influences the frequency, but which remains unknown (hidden) to researchers. Moreover, the modeling of election results and any calculation of power indices must be influenced by this hidden factor, too.

Following this idea, one can also formulate the hypothesis that the power of a decision making body as a whole and, in general, the power of each individual decision maker may depend not only on the structure of a decision making body and the distribution of votes, but also on something which may be called a hidden factor. This, at least in practice, is agreed upon by political scientists.

Index of power

Let s assume the following:

$N = \{1, ..., n\}$ denotes a set of players (these may be, e.g., individuals or parties),

ϖ_i ($i = 1, ..., n$) denotes the (real, non-negative) weight of the i-th player such that $\sum_{i \in N} \varpi_i = 1, \varpi_i \geq 0$, γ denotes a quota, γ is a real number, $0 < \gamma \leq 1$.

The (n + 1)-tuple $(\gamma, \boldsymbol{\omega}) = (\gamma, \varpi_1, \varpi_2, \dots, \varpi_n)$ such that $\sum_{i=1}^{n} \varpi_i = 1,\ \varpi_i \geq 0, 0 < \gamma \leq 1$ is called a committee of size n.

For any non-empty subset $S \subseteq N$ and a given allocation ω and quota γ, we say that $S \subseteq N$ is a winning coalition[2], if $\sum_{i \in S} \varpi_i \geq \gamma$ and a losing coalition, if $\sum_{i \in S} \varpi_i < \gamma$.

Let $G = \left\{ (\gamma, \boldsymbol{\omega}) \in \mathrm{R}_{n+1} : \sum_{i=1}^{n} \varpi_i = 1, \varpi_i \geq 0, 0 < \gamma \leq 1 \right\}$ represent the space of all committees of size n and $E = \left\{ \mathbf{e} \in \mathrm{R}_n : \sum_{i \in N} e_i = 1, e_i \geq 0 \ (i = 1, ..., n) \right\}$ a unit simplex.

By definition, a **power index** is a vector valued function

$$\pi : \mathbf{G} \to \mathbf{E},$$

where $\pi_i(\gamma, \omega)$ is the share of power that the index π grants to the i-th member of a committee.

[2] An attempt to define a power index in non-game theoretical terminology can be found in Turnovec [14], Turnovec *et al.* [15, 16].

The most commonly used definitions of such an index of power are:

- the Shapley–Shubik index of power [13],
- the Penrose–Banzhaf (absolute and relative) index of power [10], [1].

Both indices are an *a priori* measure of the power of individuals (decision makers) in relation to the formation of a coalition (Shapley–Shubik) or in terms of his/her power to break up a winning coalition (Penrose–Banzhaf) – players may be considered to be in a *swing* or *pivotal* position, respectively.

Constrained indices of power

A crucial assumption for any a priori index of power is connected with the likelihood (possibility) of a given coalition. It is assumed that any possible coalition can exist. Therefore, the power index for coalitions which are very unlikely is calculated in the same way as for coalitions which are highly likely.

The assumption about the equal likelihood of all coalitions may be contested from different points of view. For example, Rikker [11] postulates that only a minimal size (in the sense of the number of participants) coalition will be realized. Owen [9] introduced so called pre-coalitions, i.e., he defined a structural net of coalitions describing which coalitions are possible and which are not. Mercik and Mazurkiewicz [7], [8] defined the likelihood of a given coalition based on the ideological distance between members of this coalition[3], which to some extent may be seen as the beginning of estimation of the effect of hidden factors.

It is obvious that rejection of the assumption of the equal likelihood of coalitions changes the values of power indices. However, the question of whether an index of power reflects the share of power among committee members seems to have been solved already[4] – it is a common tacit supposition that a power index measures *power*, however only the absolute Banzhaf power index may be directly used as a measure of the power of a committee member[5].

---

[3] The literature about different concepts of constrained power indices is very rich. As we propose quite a different approach to solving this problem, we would like to mention the only two significant examples of such an approach: Bilbao [2] and Carreras and Freixas [3].

[4] For discussion of this dilemma, see, for example, Felsenthal and Machover [4] or Turnovec et al. [15, 16].

[5] It also seems that the Coleman power index "to act" for a decision making body may be used as a measure of power for a decision making body as a whole. In this case, however, it is not clear how to distribute *power* between members of the committee.

Any power index which assumes that coalitions are not equally likely is called a constrained power index and denoted by $\pi_i^C(\gamma,\varpi)$.

Hidden factors of power

The voting power index, $\pi_i(\gamma,\varpi)$ (or constrained voting power index $\pi_i^C(\gamma,\varpi)$) is an a priori power index and only reflects the structure of a committee. In fact, in 1953 Lloyd Shapley used his index of power as a pay-off function in a co-operative game and only later, when the theory of 0–1 (characteristic function) games was developed, did it become transformed into the Shapley–Shubik [13] power index. This a priori approach may lead to some problems in the interpretation of power indices. For example, the following two committees of 3 voters with 100 votes and a 51% majority rule: (51; 33; 34; 33) and (51; 49; 49; 2), have the same Shapley–Shubik power indices: (1/3, 1/3, 1/3). It is very hard to believe that the realization of a pay-off function for a governmental coalition (distribution of ministers, for example) will be the same for these two committees.

In the case of the three member committee (51; 49; 49; 2) one would rather expect a power index of the form $(1/3+\varepsilon,\ 1/3+\gamma,\ 1/3-\varepsilon-\gamma)$ with appropriate values for ε, γ than (1/3, 1/3, 1/3).

Hence, we consider the following equation:

$$\pi_i^R(\gamma,\varpi)=\pi_i(\gamma,\varpi)+H_i,$$

where: R stands for "realistic", $\pi_i(\gamma,\varpi)$ describes the index of power of the i-th player with a quota γ and committee of size n with structure given by ϖ and H_i stands for a hidden factor, which increases or decreases the power of the i-th player.

Of course, the hidden factor may be modeled analogically to an econometric model as

$$H_i=f(a_i^1,\dots,a_i^k)+\varepsilon_i\ ,$$

where: ε_i is the error of estimation, resulting from unknown factors influencing power[6]. We want to show that under certain assumptions H_i can be estimated for a given committee by statistical methods.

---

[6] It is quite hard to define (and maybe it is even harder to measure) the influence of a given factor on the power of a decision maker. Among such factors, one might mention: the cohesiveness of a party (measured by the probability of following the party leader), its image in the eyes of opponents or the electorate (measured, for example, by expenditure on PR). There are also more classical economic factors such as the unemployment rate, income parameters, which influence the position of a party in various coalitions, apart from the size of a party and the structure of a committee.

The following assumptions may be formulated regarding H_i for a given committee structure, voting procedure and quota:

1) for $n = 1 \Rightarrow \pi_1 > 0 \Rightarrow H_1 = 0$,

2) for $n = 2$ $\begin{cases} \pi_1 = \pi_2 > 0 \Rightarrow H_1 = 0, \\ \text{or} \\ (\pi_1 = 0 \ \ \pi_2 > 0 \ \text{ or } \ \pi_1 > 0 \ \ \pi_2 = 0) \Rightarrow H_1 = H_2 = 0 \end{cases}$

3) for $n \geq 3 \Rightarrow (\pi_i \geq 0,) \Rightarrow H_i \neq 0$ or $H_i = 0$.

It is also obvious that $\sum_{i=1}^{n} \pi_i(\gamma, \varpi) = \text{const}$ for a given committee[7].

Therefore, $\sum_{i=1}^{n} \pi_i(\gamma, \varpi) = \text{const} = \sum_{i=1}^{n} \pi_i^R(\gamma, \varpi)$ (and, of course, $= \sum_{i=1}^{n} \pi_i^C(\gamma, \varpi)$) and hence $\sum_{i=1}^{n} H_i = 0$.

Moreover, $\sum_{H_i > 0} H_i = \sum_{H_j < 0} H_j$.

We also assume that $H_i = f(a_i^1, ..., a_i^k) + \varepsilon_i$ fulfills the standard Gauss-Markov assumption, i.e.,

- $E(H_i | a_i^j) = E(H_i)$ and $D^2(H_i | a_i^j) = D^2(H_i)$,
- $\varepsilon_i \sim N(0; \sigma)$,
- $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$.

One may assume that the econometric model for H_i is linear[8]:

$$H_i = \alpha_i^0 + \alpha_i^1 a_i^1 + ... + \alpha_i^k a_i^k + \varepsilon_i,$$

where all the $a_i^1, ..., a_i^k$ are known[9] and H_i is hidden.

---

[7] This is shown for the Banzhaf-Coleman power index, for example, in Turnovec *et al*. [15]. This is obvious for standardized power indices.

[8] If not, in many cases by using simple transformations a model may be transformed into a linear one. Also, non-linear models may be used, but in this case measures of correlation which do not assume a linear relationship have to be used in the optimization problem considered.

[9] In most cases, the factors influencing various parties are different. For example, farmers' income may be used as descriptive factor for a peasant party, workers' income may be used for a socialist party, GDP may be used as a factor for both parties. In reality such factors can be astonishingly different (see, for example, Mercik, Mazurkiewicz, [8]).

The situation where $a_i^{k+1}, ..., a_i^{k+h}$ are also unknown or unmeasurable may also be considered, but in this case additional information (like the concept of "direction of influence" presented below) has to be used.

The estimation of a hidden factor

In general, the accuracy of the estimation of a hidden variable *Hi* depends on the quality of the information gained. This information may consist of the following information (see Krefft [6]):

– $\mathfrak{R}$ a vector of signs of size (1 x *k*): ($\mathcal{R}_1$, $\mathcal{R}_2$, ..., $\mathcal{R}_k$), describing the direction of influence of each *a* on the hidden variable H_i (when a sign is "+", an increase in a_i^j tends to be accompanied by an increase in H_i; when a sign is "–" an increase in a_i^j tends to be accompanied by a decrease in H_i),

– $H_i^{\max}, H_i^{\min}$ minimal and maximal values for H_i; constraints on H_i.

The estimation of $H_i^{\max}, H_i^{\min}$ is relatively easy: each H_i satisfies the following inequality:

$$-\pi_i(\gamma,\omega) \le H_i \le 1-\pi_i(\gamma,\omega).$$

The derivation of $\mathfrak{R}$ needs more than mathematical assumptions. It must be obtained by external analysis, but knowledge regarding the directions of influence is a realistic assumption.

Using these assumptions, we may define the problem of estimating H_i as the following optimization problem (OP):

Find H_i such that

$$\sum_i H_i = 0 \text{ AND}$$

{(max correlation between H_i and a_i^j if $\mathcal{R}_i$="+") OR (min correlation H_i and a_i^j if $\mathcal{R}_i$ ="–")}, for $-\pi_i(\gamma,\omega) \le H_i \le 1-\pi_i(\gamma,\omega)$.

Of course, in order to eliminate the overestimation of multi-impact on H_i the set of $a_i^1, ..., a_i^k$ must satisfy $cor(a_i^j, a_i^r) < r_\alpha$ for $j \ne r$, where *cor* stands for correlation and r_α for the critical value of correlation obtained from the *t*-Student distribution for a given *α* and *m-2* degree of freedom.

The problem of non-unique solutions

Two aspects of the uniqueness of a solution of such an OP may be considered:

– the indirect influence of correlation, which results from a given level of significance connected with the Student's *t* distribution, as well as

– directly connected to the method of solving the OP.

In both cases, the uniqueness of the set of $\{H_i\}$ is connected with the number k of the descriptive variables used to estimate it, $a_i^1, ..., a_i^k$, for $k >> 2$. It is *probably*[10] possible to find at least two different solutions when $k = 2$, but a greater number of descriptive variables excludes the possibility of such a phenomenon. Moreover, the idea of the algorithm is such that at the beginning a correlation of 1 is assumed and by appropriately lowering this correlation the first solution is found. In practice, there is no chance of finding a non-unique solution.

The additional indirect, for a given level of significance, influence of the *t*-Student critical value on the uniqueness of the set of $\{H_i\}$ is apparent. The significance level is chosen and determined by the researcher and is constant for all calculations. However, a certain parameterization may be carried out as well, in order to establish the stability of the solution. Once again, this does not produce non-unique solutions.

Example

Let us consider the situation in the Polish parliament. The a priori values of the power (Shapley–Shubik) index are presented in table 1.

Let us assume for simplicity that there are only two factors influencing power: namely factor *A* and factor *B* and that these agents are the same for all parties. Hence,

$$H_i = \alpha^i + \alpha_1^i A + \alpha_2^i B + \varepsilon .$$

Let us assume that the observations of *A* and *B* are known (table 2). The data are taken from Gładysz and Mercik (2004) and describe the rate of unemployment and the voting frequency during the presidential election in a random sample based on 20 of the 316 electoral districts. The correlation coefficient between *A* and *B* equals –0.442 and is not significant for $\alpha = 0.05$ and $df = 18$. Therefore, *A* and *B* are not signifi-

[10] In fact, an attempt to find such a counter-example which fulfills the condition $\sum_i H_i = 0$ was unsuccessful.

cantly correlated. Moreover, they satisfy the Gauss–Markov assumptions for econometric models.

Table 1. The Shapley–Shubik power index for the lower chamber of the Polish parliament (as of June 6, 2006)

Party	Number of seats	Shapley–Shubik power index	Penrose–Banzhaf power index	Holler–Packel power index
PiS	155	36.67	35.71	21.43
PO	131	23.33	21.43	14.29
Sam	55	13.33	14.29	17.86
SLD	55	13.33	14.29	17.86
LPR	29	6.67	7.14	14.29
PSL	25	6.67	7.14	14.29
AKP	5	0	0	0
N/A	5	0	0	0

Table 2. Data for factors A and B, where a_i and b_i are their respective observations

a_i	b_i
0.746047316791691	0.541744093443058
0.491160748947353	0.607071978314536
0.690007108225384	0.661037636476116
0.76255433494033	0.56967550274223
0.790111694726154	0.591724497883281
0.721205754471972	0.615664433392248
0.66677645027938	0.671394146754226
0.723926380368098	0.561817788825246
0.748810162220137	0.595096300050684
0.75118025751073	0.615207602380716
0.679986921931195	0.585003135545137
0.754705236177224	0.55968544753041
0.695804841988542	0.575310389893269
0.752734972356193	0.603793565266141
0.672882071488883	0.590124883714374
0.793553149606299	0.576263097831725
0.770094914943372	0.569689921210831
0.679386257505003	0.593315329979284
0.729782097706626	0.585657059619186
0.620696911555363	0.67195909848916

In order to estimate (for the PiS party in the case of the Shapley–Shubik power index) the value of H_1, one has to carry out the following steps:

– define the possible range of H_1:

• $\pi_1(\gamma,\omega) \le H_1 \le 1-\pi_1(\gamma,\omega) \Rightarrow -0.3667 \le H_1 < +0.6333$,

• find a solution of the OP: some of the calculations are presented in table 3, where a macro written for the Excel package was used to generate (hidden) values of H_1,

• calculate the value of H_1 for given values of a_i and b_i. In this example a_1 and b_1 are 0.746047316791691 and 0.541744093443058 respectively.

Table 3. The results of the calculations for the hidden H_i

H_i from solver	a_i	b_i
-0.153627944542541	0.746047316791691	0.541744093443058
-0.146757947873622	0.491160748947353	0.607071978314536
-0.153143714020534	0.690007108225384	0.661037636476116
-0.152415087978025	0.76255433494033	0.56967550274223
-0.15631441031088	0.790111694726154	0.591724497883281
-0.152950254207545	0.721205754471972	0.615664433392248
-0.153480022195008	0.66677645027938	0.671394146754226
0.113387718418366	0.723926380368098	0.561817788825246
0.113773865925809	0.748810162220137	0.595096300050684
-0.0197994396451609	0.75118025751073	0.615207602380716
-0.0448218391859075	0.679986921931195	0.585003135545137
-0.137963041157895	0.754705236177224	0.55968544753041
0.021849590380696	0.695804841988542	0.575310389893269
-0.000719861564807733	0.752734972356193	0.603793565266141
0.413318880232058	0.672882071488883	0.590124883714374
0.171405580331966	0.793553149606299	0.576263097831725
0.085864414827165	0.770094914943372	0.569689921210831
-0.0464718919988504	0.679386257505003	0.593315329979284
0.104089532745811	0.729782097706626	0.585657059619186
0.288119880519647	0.620696911555363	0.67195909848916

The same steps have to be carried out for the remaining H_i. It is crucial that the condition $\sum_i H_i = 0$ be satisfied. However, the estimated values of H_i do not satisfy it! Once again, one can use the Excel macro with elements of annealing[11] to obtain an optimal (or near-optimal solution). The result for this example is presented in table 4.

Due to the choice of parameters (our goal was and not only to illustrate the methodology of estimating the power of certain group members than to estimate the power of parties represented in the lower chamber of the Polish parliament), the differences obtained in the example are minimal. Moreover, in this case the obtained solution is only near-optimal: $\sum_i H_i = 0.001520265757$.

[11] Piotr Wawrzynowski, a PhD student from Wroclaw University of Technology, wrote the appropriate program in VBA.

Table 4. The values of H_i and the corrected Shapley–Shubik index for the example considered

No.	Party	Shapley–Shubik power index	Values of standardized H_i	Corrected Shapley–Shubik power index
1	PiS	36.67	–0.0594209	36.6105791
2	PO	23.33	1.0667074	24.3967074
3	Sam	13.33	–0.0909267	13.2390733
4	SLD	13.33	–0.2178984	13.1121016
5	LPR	6.67	–0.417085	6.252915
6	PSL	6.67	–0.1293497	6.5406503
7	AKP	0	7.13523E-11	7.13523E-11
8	N/A	0	4.86937E-12	4.86937E-12

Conclusions

Hidden factors exist which influence power indices. In each problem the set of hidden factors is different and has to be determined by specialists (most likely political scientists). The proposed method for deriving the influence of these factors is effective and gives corrected values of any power index (not just the Shapley-Shubik index). It is expected that the corrected distribution of power is a better reflection of the real distribution of power among decision-makers (voters) in a given decision-making body.

Bibliography

[1] BANZHAF J.F., *Weighted Voting Doesn't Work: A Mathematical Analysis*, Rutgers Law Review, 1965, 19, s. 317–343.

[2] BILBAO J.M., *Axioms for the Shapley value on convex geometries*, European Journal of Operational Research, 1998, 110, s. 368–376.

[3] CARRERAS F., FREIXAS J., *Semivalues versatility and application*, Annals of Operations Research, 2002, s. 343–358.

[4] FELSENTHAL D.S., MACHOVER M., *A Priori Voting Power: What is it About?* Political Studies Review, Vol. 2, issue 1, 2004, s. 1–23.

[5] GLADYSZ B., MERCIK J.W., *Econometric modeling*, Wrocław University of Technology Press, Wroclaw 2004 (in Polish).

[6] KREFFT A., *Diagnostic function of unobserved phenomenon*, Wrocław University of Technology Press, Wrocław 1999 (in Polish).

[7] MAZURKIEWICZ M., MERCIK J., *Modified Shapley–Shubik power index for parliamentary coalitions*, Operational Research and Decisions, 2005, 2/2005, s. 43–52.

[8] MERCIK J., MAZURKIEWICZ M., *Modelling the behaviour of Polish members of parliament*, Operational Research and Decisions, 2001, 2/2001, s. 77–90.

[9] OWEN G., *Values of games with a priori unions*, [in:] *Mathematical Economy and Game Theory*. Springer Verlag, Berlin 1977.
[10] PENROSE L.S., *The Elementary Statistics of Majority Voting*, Journal of the Royal Statistical Society, 1946, 109, s. 53–57.
[11] RIKKER W.H., *The Theory of Political Coalitions*, Yale University Press, New Haven, 1962.
[12] SHAPLEY L.S., *A Value for n-Person Games*, [in:] Kuhn H.W. and Tucker A.W. (eds.), *Contributions to the Theory of Games II*, Annals of Mathematical Studies, 28. Princeton University Press. Princeton, 1953, s. 307–317.
[13] SHAPLEY L.S., SHUBIK M., *A Method for Evaluating the Distribution of Power in a Committee System*, American Political Science Review, 1954, 48 (3), s. 787–792.
[14] TURNOVEC F., *New Measure of Voting Power*, Czech Economic Review, Acta Universitatis Carolinae Oeconomica, 2007, 1, s. 1–11.
[15] TURNOVEC F., MERCIK J., MAZURKIEWICZ M., *I-Power and P-Power: Shapley-Shubik AND/OR Penrose-Banzhaf?* [in:] *Operational Research and System Science* (R. Kulikowski. J. Kacprzyk and R. Slowinski – eds.), Warszawa, 2004, s. 121–127.
[16] TURNOVEC F., MERCIK J., MAZURKIEWICZ M., Power indices: A unified approach. *Paper presented at European Public Choice Society Annual Meeting*. Durham 2004.
[17] TURNOVEC F., MERCIK J., MAZURKIEWICZ M., *Power indices methodology: Decisiveness, Pivots and Swings*, To be published in Special Edition in honor of Manfred Holler by Springer Verlag, 2007.

Estymacja ekonometryczna czynników ukrytych w decyzjach grupowych – wpływ na estymację indeksu siły

Estymacja siły decydenta jest zależna od właściwej oceny wszystkich czynników na nią wpływających. Budowane zazwyczaj modele ekonometryczne są co prawda statystycznie istotne, jednak mają relatywnie niską determinację. W artykule zaproponowano ekonometryczną technikę modelowania, służącą do podniesienia wartości współczynnika determinacji i w konsekwencji prowadzącą do poprawy oceny siły decydenta. Takie podejście umożliwia także zbadanie charakterystyk różnych czynników ukrytych dzięki obserwacji ich oddziaływania na zmienność modelu. W pracy pokazano także przykład oraz algorytm takiego postępowania w podejściu do modelu ekonometrycznego.

Słowa kluczowe: *zmienne ukryte, indeks siły*