WP 22
23 April 2010

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

STATISTICAL OFFICE OF THE EUROPEAN UNION (EUROSTAT)

CONFERENCE OF EUROPEAN STATISTICIANS

Joint Eurostat/UNECE Work Session on Demographic Projections
(28-30 April 2010, Lisbon, Portugal)

Item 9 – Forecasting demographic components: migration

Dealing with uncertainty in international migration predictions: from probabilistic forecasting to decision analysis

Jakub Bijak[1], University of Southampton, United Kingdom

1. INTRODUCTION

The paper focuses on the uncertainty of international migration predictions, as well as on their consequences for population projections and information delivered to decision makers. One of the key questions about probabilistic population forecasting is how its outcomes – the predictive distributions – can be useful for policy making and planning purposes. Some insights in that respect can be drawn from the statistical decision analysis, which takes into account the potential costs of both under- and overestimation of the variables under study, for example of current or future migration flows. The on-going paradigm shift in demographic projections, from deterministic to stochastic, can thus be brought even further, to the field of decision support. In that regard, the paper presents the preliminaries of Bayesian decision analysis together with some examples concerning international migration forecasts.

Another important issue concerns the assumptions made about the migration component of demographic projections. In particular, the consequences of various assumptions concerning stationarity and variability of migration processes are discussed and evaluated. In this context, the limitations of predictability are sketched for consideration both by forecast providers and users. Such limitations include, among others, the plausible horizon of population predictions, as well as realistic expectations with respect to their outcomes. In addition, several decision-making strategies under low predictability, alternative to formal statistical analysis, are discussed on the basis of recent advancements in the decision sciences. It is argued that instead of striving for unachievable precision of forecasts, especially with respect to migration, the providers and users of demographic predictions could make a joint attempt to encompass the inevitable uncertainty within the decision-making process. With focus on this objective, a draft outline of interactive population forecasting based on the Bayesian statistical paradigm is proposed, together with a brief discussion of some promising areas of future research.

The current paper makes an attempt to deal with the uncertain forecasts from the policy-oriented perspective of forecast users (decision makers). Hence, Section 2 presents a brief introduction to the decision analysis from the Bayesian perspective. Section 3 contains an overview of literature and discussion on the generic limits of predictions under uncertainty from the point of view of forecast users. Finally, in Section 4, an interactive approach to demographic forecasting is proposed, with an increased role of the dialogue between forecasters and users.

[1] Centre for Population Change, Division of Social Statistics and S3RI, School of Social Sciences, University of Southampton, Southampton SO17 1BJ, United Kingdom. Email: J.Bijak@soton.ac.uk

This research is funded by ESRC Grant number RES-625-28-0001. The Centre for Population Change is a joint initiative between the University of Southampton and a consortium of Scottish Universities in partnership with ONS and GROS. The findings, interpretations, and conclusions expressed in this paper are entirely those of the author and should not be attributed in any manner to ONS or GROS.

An extended version of this paper is to appear in Chapter 11 of J. Bijak (2010) *Forecasting International Migration in Europe: A Bayesian View*. Springer Series on Demographic Methods and Population Analysis; Springer, Dordrecht.

2. INTRODUCTION TO BAYESIAN DECISION ANALYSIS

Decision analysis and support cover many methods, such as mathematical statistics and operational research, and many areas of applications. In public policy and planning related to demography and migration, decisions are made on a variety of levels, from local authorities through national governments to multi-national bodies, such as the European Union. A distinction can be made here between planning-related decisions, as in the case of spatial organisation, and more general and strategic policy design, such as immigration policies of particular countries or of the whole EU. Planning-related decisions require some numerical input, whereas the strategic ones are guided by more general, largely qualitative, advice; still, all of them are made under the conditions of uncertainty. This section deals with a statistical decision analysis, suitable for the former (quantifiable) types of problems, while Section 3 provides some general guidance to the second type of policy challenges, proposed on the basis of recent advancement in forecasting and decision science.

The approach presented here follows the Bayesian perspective, since the very axiomatic construction of Bayesian statistics is firmly grounded in the decision analysis, as thoroughly discussed e.g. by DeGroot (1970/1981), Bernardo and Smith (2000), or Robert (2001). Within the framework of statistical decision problems, such as estimation or prediction, the decisions often concern the choice of one value from the relevant probability distributions, depicting the possible 'states of nature' and how likely are they to occur. In the context of demography, the decision analysis was discussed for example by Alho and Spencer (2005), referring to population estimates, whereas the current paper will focus more on applications to population forecasting.

The decision-analytic foundations of Bayesian statistics rely on the axiomatic definition of the (usually bounded) utility function u, being a measure of preference, defined over the space of possible outcomes $\mathbf{\Omega}$ of decisions $\mathbf{D}$, so that $u\colon \mathbf{\Omega} \times \mathbf{D} \to \mathbf{R}$ (*idem*). As most persons, and even more so the policy makers, appear to be risk-averse, their utility functions are concave and bounded from above, convex functions being reserved for "risk lovers" (Robert, 2001: 59). For public policy applications the cautious attitude to uncertainty is especially vital due to the possible large-scale consequences of wrong decisions, in particular involving taxpayers' money.

Within the Bayesian paradigm, the decision analysis offers a useful framework for calculating point estimates or forecasts from the relevant posterior or predictive distributions. In order to do it, a non-negative *loss function* $L\colon \mathbf{\Omega} \times \mathbf{D} \to \mathbf{R}$, has to be defined over the space of possible states of nature[2], $w \in \mathbf{\Omega}$, and decisions $d \in \mathbf{D}$. The loss function can be defined simply as negative utility, $L(w, d) = -u(w, d)$ (DeGroot, 1970/1981: 106; Robert, 2001: 60). This function describes the loss (or cost) of making particular decisions d about w, in particular, the wrong ones.

In order to obtain an optimal decisions in a Bayesian framework, let P denote the probability distribution defined over $w \in \mathbf{\Omega}$. For every possible decision $d \in \mathbf{D}$, the expected loss under this distribution, known as *risk* and denoted by $\rho(P, d)$ can be calculated as (e.g. DeGroot, 1970/1981: 106):

(1a) $$\rho(P, d) = \int_{\Omega} L(w,d)\,p(w)\,dw \text{, for continuous } P \text{, or:}$$

(1b) $$\rho(P, d) = \sum_{w \in \Omega} L(w,d)\,p(w) \text{, for discrete } P,$$

where $p(w)$ respectively denotes the density or probability function of the distribution P. It is additionally assumed that such expected loss exists and is finite. The optimal Bayesian decision d^* is then such d for which the risk (1) is minimised (*idem*):

(2) $$d^* = \arg\min_{\{d\}}\{\rho(P, d)\}.$$

This decision is specific to the probability distribution P and the loss function L. In practice, P is usually either a posterior distribution for estimation problems, or a predictive distribution in forecasting applications.

One of the important issues concerning the choice of a loss function is its symmetry. Symmetric functions can be useful to obtain point estimates of the central characteristics of the relevant distributions. It can be shown (e.g., Bernardo and Smith, 2000: 257) that the quadratic loss function $L(w, d) = a\,(w - d)^2$ yields the mean of the posterior or predictive distribution as the optimal solution, whereas the absolute value function $L(w, d) = a\,|w - d|$ yields the median. Similarly, for the point function $L(w, d) = 1 - \mathbf{1}_{w=d}$, where $\mathbf{1}_X$ is the indicator function, equal one if X holds and zero otherwise, the optimal solution is the mode of the distribution. In all cases it is assumed that the relevant characteristics exist (*idem*). However, as noted by Lawrence *et al.* (2006), in many real-life forecasting situations, loss function is asymmetric.

A simple example of an asymmetric loss function is of the linear-linear (LinLin) form, whereby (e.g. Bernardo and Smith, 2000: 257):

[2] In the general case $\mathbf{\Omega}$ is not restricted to the space of model parameters $\mathbf{\Theta}$, but can depict any quantity of interest, such as forecasts, in all instances reflected through the respective posterior or predictive distributions depicting uncertainty. Notation in this paper follows DeGroot (1970/1981).

(3) $$L(w, d) = a \cdot (w - d) \cdot \mathbf{1}_{w \leq d} + b \cdot (d - w) \cdot \mathbf{1}_{w > d}.$$

The optimal solution of the decision problem is then the quantile of rank $b/(a+b)$ from the relevant distribution (*idem*). For $a = b$ the problem thus reduces to the one under the absolute value function $a\,|w - d|$, and yields a median solution, as discussed above. For more complex, possibly non-linear cases, Varian (1975) introduced the linear-exponential (LinEx) loss function, defined as (after: Zellner, 1986: 446):

(4) $$L(w, d) = b \cdot \{\exp[a \cdot (w - d)] - a \cdot (w - d) - 1\}.$$

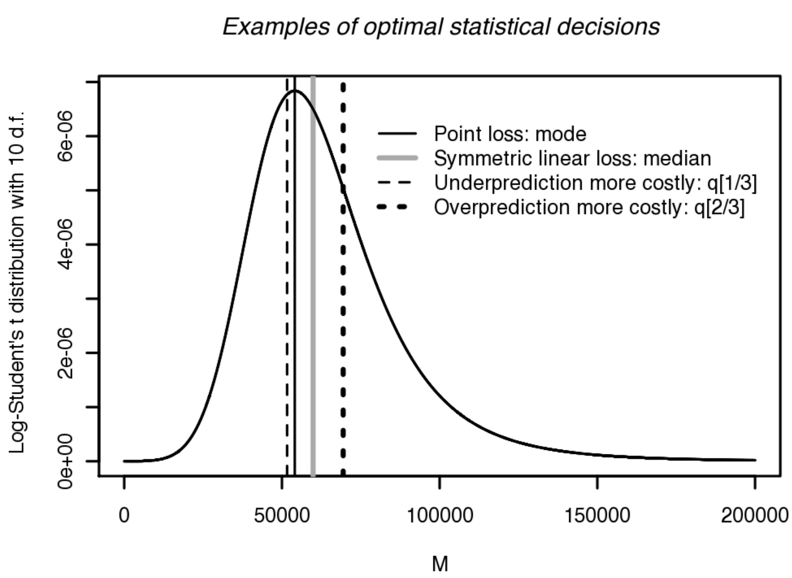
This function is "almost linear" for $w < d$ and "almost exponential" for $w > d$ in cases when $a > 0$, and the opposite holds when $a < 0$ (*idem*). It can be also shown (Zellner, *idem*: 447) that the optimal decision is $d^* = (-1/a) \cdot \ln\{E_w[\exp(-aw)]\}$, where $E_w(\cdot)$ is the expected value with respect to the distribution of w over $\mathbf{\Omega}$. It is assumed that this expected value exists, as for example in the case of Normal distributions. Noteworthy, $E_w[\exp(-aw)]$ is the moment-generating function for the density of w (*idem*).

As an example, consider a Bayesian forecast of registered immigration based on relatively low-precision assumptions a priori, short time series of data and a simple random-walk model (see e.g. Bijak and Wiśniowski, 2010). Let the forecasted number of immigrants, M, follow a predictive log-t distribution such that $3 \cdot [\ln(M) - 11] \sim t_{10}$, where t_{10} denotes a central t distribution with 10 degrees of freedom, mean zero and precision of one. For linear (LinLin) loss functions, the appropriate decisions are quantile-based and thus invariant under positively monotonous transformations. The appropriate quantiles q can be computed by transforming the relevant quantiles from the t_{10} distribution, q', using the formula $q = \exp(q'/3+11)$.

Hence, for instance, where underestimation is twice as costly as overestimation, the decision would be based on the upper tertile of the predictive distribution. In this case, the point forecast of the number of immigrants equals about 69,448. In the opposite situation, with overestimation twice more costly, decision is based on the lower tertile, that is, 51,620 immigrants. In turn, for the symmetric linear loss function the optimal choice is the predictive median (59,874 immigrants). Finally, under a point loss, the resulting decision is the mode of the predictive distribution, usually not invariant under transformations, which has to be derived numerically from the target log-t distribution (the solution being about 54 thousand immigrants). All the decisions from the presented stylised examples are shown in Figure 1 together with the underlying predictive distribution. It is worth noting that if the loss functions were of higher orders, for example quadratic or LinEx, optimal decisions would not exist, since due to the heavy tails, the log-t distributions do not have positive moments.

From the statistical point of view, the quantile-based solutions under simple LinLin functions have convenient properties. Firstly, quantiles are robust against the presence of outliers in the distributions. Secondly, they remain invariant under positively monotonous transformations. Thirdly, alternative statistics may not always exist, as moments in heavy-tailed distributions, or more than one solution may exist, as distributions can have many modes. Besides, from the point of view of interactions between forecasters and forecast users, elaborated further in Section 3, such functions as the LinLin can be more straightforward to elicit from the forecast users.

Figure 1. Examples of optimal decisions for a log-*t* distribution with 10 d.f.



Source: own elaboration in R

3. LIMITATIONS OF MIGRATION AND POPULATION PREDICTIONS

The optimal decisions, such as the ones outlined in the precious section, can be useful in the context of planning, where the decisions should be based on certain numerical parameters, such as the expected numbers of migrants, associated costs, etc. However, there are many other decision settings in which, given the prevalent uncertainty, the room for manoeuvre is much more limited. In that regard, the most notable problems with the traditional approach to forecasting include assumptions that the future will resemble the past and that events under study are independent, which need not hold in complex, network-based systems (Makridakis and Taleb, 2009). Among other flaws of the traditional approach, the use of tractable, thin-tailed error terms (such as Gaussian), the assumption of the existence of a finite variance[3], and the human tendency to underestimate uncertainty can be mentioned (*idem*).

A synthetic typology of decision situations under uncertainty was proposed by Taleb (2009). He distinguished four classes according to the type of uncertainty (thin-tailed versus all other types, including unknown); and the type of payoffs (linear versus non-linear). The 'payoffs' here are akin to the utility (or negative loss) functions introduced in Section 2. Taleb's analysis further focused on the hardly- or completely unpredictable events carrying possibly non-linear payoffs. As discussed in the previous section, for the thin-tailed (e.g. Gaussian) uncertainty, optimal decisions exist even for some non-linear loss functions, such as LinEx. Under linear loss functions, moment-based optimal decisions exist even for heavy-tailed distributions, assuming that the latter can be properly approximated. However, in heavy-tailed non-linear cases, the statistical decision theory fails. In such instances, it has to be replaced by general, common-sense decision-making strategies, unless the problem can be reduced to other types, for example by changing or bounding the loss function (*idem*)[4]. This is important especially in the migration context, where the relevant distributions can rarely be expected to be thin-tailed, due to the dynamic and changing nature of the process.

Makridakis and Taleb (2009) summarised several common-sense strategies for the use of forecasts. Their most important recommendations include: avoiding the "illusion of control" (or illusion of having accurate predictions, which can bring about dangerous consequences; also relevant for demographic and migration applications), adopting protective strategies, and setting up backup plans and additional "reserves" of resources (*idem*). Such reserves may seem redundant and unnecessary from the point of view of optimal decision making, although the latter, labelled by Taleb (2009) as "overoptimisation", was heavily criticised for making the complex systems in question much more vulnerable to unpredictable or hardly predictable events.

Another strategy suggested by Makridakis and Taleb (2009) was to apply the "minimax" approach to decision making, minimising the maximum potential losses (in the authors' terminology: maximin, maximising the minimal payoffs). However, from the Bayesian point of view this strategy, if it uniquely exists, has several drawbacks. Minimax decisions exhibit bias towards the worst-case scenarios (the least favourable prior distributions), do not take into account all information available and despite their construction can sometimes lead to worse outcomes than the approaches that are less pessimistic with respect to the states of nature (Robert, 2001: 66–77). According to Bernardo and Smith (2000: 449), although some minimax solutions may be acceptable as optimal Bayesian decisions under certain pessimistic priors, a general minimax rule "seems entirely unreasonable". Besides, in practical application, derivation of the least favourable distributions may pose a serious problem. Attempts to do so include the analysis of robustness of Bayesian decisions against changes in prior distributions. Some options here consist in limiting the optimisation of the risk function $\rho(p, d)$ to a certain class $\mathbf{\Gamma}$ of prior distributions p. The resulting solutions are referred to as *conditional* $\boldsymbol{\Gamma}$*-minimax* estimates or predictions (Męczarski, 1998).

The notion of conditional Γ-minimax decision rules has led to a concept of 'stable' estimates or predictions. As defined by Męczarski (1998: 113), a *stable decision* $d^{\#}$ with respect to a parameter θ or prediction x^P is the one, for which the oscillations of risk $\rho(P, d)$ are minimal for all prior probability distributions $P \in \mathbf{\Gamma}$:

(5) $\sup_{\{P\in\mathbf{\Gamma}\}}\{\rho(P, d^{\#})\} - \inf_{\{P\in\mathbf{\Gamma}\}}\{\rho(P, d^{\#})\} = \inf_{\{d\in\mathbf{D}\}}\{\sup_{\{P\in\mathbf{\Gamma}\}}\{\rho(P, d)\} - \inf_{\{P\in\mathbf{\Gamma}\}}\{\rho(P, d)\}\}.$

Męczarski (1998) offered a number of analytical solutions for $\mathbf{\Gamma}$-minimax and stable decisions $d^{\#}$ for some classes of prior distributions, whilst noting that a more general treatment of different statistical models seems hardly possible. Nevertheless, from a policy perspective, the presented notions are certainly appealing, since they could potentially inform the policy makers, how robust are their decisions against different types of uncertainty depicted by prior distributions p. For demographers, exploring these options would additionally enrich the possibilities offered by decision analysis to the practical applications, following the suggestions of Alho and Spencer (2005).

Regardless of the future methodological advancements, from the point of view of forecast users, a crucial question becomes: what types of decision problems can be answered or aided by forecasts. In this context, Orrell (2007) argued that appropriate risk assessments are crucial, especially if the potential dangers (negative payoffs) are large. At the same time, Orrell (2007) warned against being too risk averse, which can lead to negative externalities in such situations,

[3] Note that in the examples presented in the current study, the log-*t* predictive distributions for migration rates are heavy-tailed and their positive moments (including variance) do not exist.

[4] As one of the practical ways of putting 'caps' on payoffs or loses, Taleb (2009) proposed insurance, although admitting that this strategy may not work well under very heavy tails, such as in case of catastrophe insurance (and reinsurance).

when policy responses can be more damaging than the problem they were trying to resolve. In the context of migration, examples of such externalities of either too restrictive or too lax immigration policies can be, respectively, the loss of human capital of potential migrants, or an increased financial strain on public services and possible challenges to social cohesion.

One reason for extreme responses, as suggested by Lawrence *et al.* (2006: 504), can be that forecast users tend to focus on extreme probabilities of events, close to zero or one, and thus prefer such $1 - \gamma$ predictive intervals, for which γ is very small (for example, $1 - \gamma = 0.95$ or 0.99). This is an additional argument for presenting predictive intervals for lower probabilities (e.g., with $1 - \gamma = 0.8$ or 0.667), as a way to avoid the illusion of control (see also Lutz *et al.*, 2004: 37). Lawrence *et al.* (2006) cite several studies which suggest that, notwithstanding, forecast users tend to prefer interval forecasts to point forecasts, the former clearly providing more information.

Therefore, a tentative recommendation for the forecasters and the forecast users would be that interval forecasts are useful and can provide valuable information for the decision making, although the intervals should not be based on too high probabilities in order to avoid overconfidence. Narrower probability ranges suggest additional caution, as the probability that the variables under study fall outside the predictive intervals cannot be in such cases seen as negligible and ignored. On the part of the users, as noted by Lawrence *et al.* (2006), an additional caveat would be that the performance and expertise of the forecasters should not be assessed on the basis of their ability to minimise the width of interval whilst maximising the probability. Such forecasts are not only very likely to miss, but also to contribute to unjustified "illusion of control" among the policy makers and ultimately generate further problems in addition to the ones they were supposed to contribute to solving.

A migration-related example of very narrow predictive intervals is the forecast of post-EU enlargement immigration to the United Kingdom (Dustmann *et al.*, 2003). The underprediction of actual flows by over one order of magnitude resulted, among others, from assuming stationarity of the underlying process, which assumption in case of migration can be problematic. At this point it is worth reiterating the potential of Bayesian methods, which allow for including expert judgement, for example on the low precision of forecasts, next to the data. Moreover, the Bayesian interpretation of probabilities as subjective measures of belief, if made explicit to the users, can be also helpful in avoiding overconfidence in forecasts and admitting their inherent frailties. Finally, the limited predictability of such volatile processes as migration poses limits on plausible forecast horizons. For example, the expert-based Bayesian forecasts of immigration into seven European countries, prepared by Bijak and Wiśniowski (2010) suggest horizons of ten years at most, echoing earlier suggestions of Holzer (1959).

4. CONCLUSION: FROM PREDICTIONS TO DECISIONS

As argued before, the Bayesian approach can provide an umbrella framework for forecasting and decision making, providing a coherent mechanism of inference and decision support. However, unlike in other approaches to forecasting, the decision support requires a dialogue between forecasters and decision makers aimed at tackling a specific decision problem. This dialogue can further include experts in the field, who can provide prior information, especially vital in the absence of reliable quantitative data, as it is often the case in migration studies. Besides, as argued by Lawrence *et al.* (2006), combining judgement with data, the very essence of Bayesian inference, leads to better forecasts than relying on either data or judgement alone.

Such interactive expert-based Bayesian forecast of migration (or population) could be summarised as follows. After the forecast users have formulated the problem, the decision framework is elicited from them by researchers (forecasters). This framework includes the loss functions, required horizon and other parameters of the decision. In order to make full use of the possibilities offered by the Bayesian approach, prior distributions of the parameters of the forecasting models can be elicited from the domain experts. Subsequently, these elements are then combined with data in the forecasting models, and the final outcomes – forecasts – are reported back to the users.

The final outcome of the procedure is a set of user-specific forecasts enhanced by simple decision advice provided to the decision makers. Such forecasts should ideally comprise other elements and caveats, most importantly including an explicit uncertainty assessment and a clear statement of the limits of predictability. Such interactive forecasts would inevitably lose generality, having to respond to specific problems faced by the decision makers. The researchers would also be no longer fully autonomous in preparing the forecasts and interpreting their outcomes, as these would emerge in a multi-stage process involving dialogue with forecast users and possibly also other experts. In this way, the paradigm shift in demographic forecasting from deterministic point forecasts, through variant to stochastic predictions, would continue towards the decision-analytic outcomes. Such forecasts would then become an explicit tool of well-defined decision support rather than merely a numerical exercise.

Also the distinction between specific planning-related and more general policy-relevant decisions will have an impact on what is possible in terms of decision support from the point of view of the forecasters. In the former case it can be a proper statistical decision analysis, such as the Bayesian one presented before, while in the latter it can be, for example, a set of scenarios, equipped with clear caveats about uncertainty. In the same way as the forecasters should not promise the impossible and clearly state the limits of predictability, the users should not expect the impossible from the providers of predictions. Therefore, confronting the users' expectations with what is actually possible from the scientific point of view should constitute the most important element of the dialogue between the forecasters and forecast users.

5. REFERENCES

Alho, J., and Spencer, B. (2005). *Statistical Demography and Forecasting*. Springer, Berlin-Heidelberg.

Bernardo, J. M., and Smith, A. F. M. (2000). *Bayesian Theory*. John Wiley, Chichester.

Bijak, J., and Wiśniowski, A. (forthcoming 2010). Bayesian forecasting of immigration to selected European countries by using expert knowledge. *Journal of the Royal Statistical Society* A, 173.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York [(1981). *Optymalne decyzje statystyczne*. PWN, Warszawa].

Dustmann, C., Casanova, M., Fertig, M., Preston, I., and Schmidt, C. M. (2003). *The impact of EU enlargement on migration flows*. Home Office Report 25/03. Home Office, London.

Holzer, J. Z. (1959). *Prognoza demograficzna Polski na lata 1960–1975 wg województw* [*Demographic forecast of Poland for 1960–1975, by voivodships*]. PWG, Warsaw.

Lawrence, M., Goodwin, P., O'Connor, M., and Önkal, D. (2006). Judgemental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, 22(3): 493–518.

Lutz, W., Sanderson, W. C., and Scherbov, S. (eds.) (2004). *The End of World Population Growth in the 21st Century: New Challenges for Human Capital Formation and Sustainable Development*. Earthscan, London.

Makridakis, S., and Taleb, N. (2009). Living in a world of low levels of predictability. *International Journal of Forecasting*, 25(4): 840–844.

Męczarski, M. (1998). *Problemy odporności w bayesowskiej analizie statystycznej* [*Robustness issues in Bayesian statistical analysis*]. "Monografie i opracowania" no. 446, Warsaw School of Economics, Warsaw.

Orrell, D. (2007). *The Future of Everything. The Science of Prediction*. Thunder's Mouth Press, New York.

Taleb, N. N. (2009). Errors, Robustness, and The Fourth Quadrant. *International Journal of Forecasting*, 25(4): 744–759.

Varian, H. R. (1975). A Bayesian Approach to Real Estate Assessment. In: S. E. Fienberg and A. Zellner (eds.), *Studies in Bayesian Econometrics and Statistics. Essays in Honor of Leonard J. Savage*. North-Holland, Amsterdam: 195–208.

Zellner, A. (1986). Bayesian estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Association*, 81(394), 446–451.

■