# Experts adjusting model-based forecasts and the Law of Small Numbers

Philip Hans Franses

*Econometric Institute*

*Erasmus University Rotterdam*

Econometric Institute Report 2007-42

**Abstract**

This paper conjectures that the behaviour of experts who adjust statistical-model-based forecasts obeys the Law of Small Numbers [LSN]. To put this hypothesis to an empirical test, I propose a simple but effective methodology. It is applied to a database containing information on many experts and their adjustments to forecasts of monthly sales of a variety of products. I find strong evidence in support of the LSN. Implications of this finding are discussed.

Key words: Law of Small Numbers; Expert adjustment; Model-based forecasts

October 2 2007

Address for correspondence: Econometric Institute, Erasmus University Rotterdam, P.O.Box 1738, NL-3000 DR Rotterdam, The Netherlands, franses@few.eur.nl. I am very grateful to Rianne Legerstee for excellent research assistance and to Sander Demouge and colleagues at NV Organon (Oss, The Netherlands) who allowed me to use their extensive database for academic research.

## 1. Introduction and the main hypothesis

Statistical models are frequently used for out-of-sample forecasting. Such models have as input recent values of explanatory variables, parameter values are updated, and the models are extrapolated into the future. To approximately indicate the two ends of the spectrum, consider macro-economic models which may contain hundreds of equations, aiming to cover all aspects of an economy, and single-equation exponential smoothing models designed to predict future sales. By now, modellers can choose from a range of models and from a range of parameter estimation methods, while computer power and data quality are getting better, as we speak.

Even though progress has been and will be made in statistical modelling skills and techniques, it is rare to see that model-based forecasts are blindly incorporated in budgeting and allocation plans. Indeed, quite often an expert with domain knowledge modifies the model-based forecast by changing direction or value. Macro-economic model-based forecasts for, say, inflation might be adjusted based on recent events in adjacent countries, and monthly pharmaceutical sales forecasts are adapted if new legislation is foreseen to be activated. This experts' touch is widely considered as normal and useful, and there is a literature that suggests that model modifications can be beneficial to forecast accuracy. In this paper I examine whether this experts' touch mimics characteristics that occur in related decision-making settings.

### Experts adjusting model-based forecasts

To focus the discussion, consider a rather stylized situation, which does fit well with the empirical analysis below, but which can also easily be extended to more complicated settings. Consider the following variables

| | |
|---|---|
| $MF_{t+h\|t}$ | $h$-step-ahead model-based forecast (made from origin $t$) |
| $EF_{t+h\|t}$ | $h$-step-ahead expert forecast (made from origin $t$) |
| $Y_{t+h}$ | realization at time $h$ |

The nature of the model does not matter, although I assume that it is some kind of regression-based model. This means that the model forecasts have a tendency to convert to the mean value of the variable $Y$ when $h$ gets larger. A typical situation is that an expert receives the model-based forecast *MF*, evaluates its merits, and eventually possibly decides to modify it towards *EF*. Note that the expert and the model-builder can be one and the same person.

Suppose now that the model-based forecast is a linear autoregressive function of past data, that is,

$$MF_{t+h|t} = \mu_1 + \rho_1 Y_t + \rho_2 Y_{t-1} + \rho_3 Y_{t-2} + ..... \tag{1}$$

Here the underlying model is an autoregressive model for *Y,* and the parameters in forecasting scheme (1) are functions of the parameters in this model. It is quite common to apply a recursive forecasting procedure, which means that parameters are estimated for, say, *R* in-sample data, and then one-step-ahead to *h*-steps-ahead forecasts are made. Next, the sample is increased to *R*+1, parameters are updated and again one-step-ahead to *h*-steps-ahead forecasts are made. Denote the number of forecasts as *P*.

The expert receives the model-based forecasts and makes an adjustment. The expert has observed current and past *Y* as well. Also, he or she may have information on *k* domain-specific variables captured in *X* which can be extrapolated by the expert and which he or she deems relevant. So, an expert-adjusted forecast can be written as

$$EF_{t+h|t} = \mu_2 + \delta_1 Y_t + \delta_2 Y_{t-1} + \delta_3 Y_{t-2} + ..... + \beta X_t + ... \tag{2}$$

where $\beta$ is an (1 x *k*) vector of parameters fixed by the expert.

To my knowledge, it is very uncommon for experts to write down explicitly what is in *X*, how he or she gets values for $\beta$, and whether the $\delta$ parameters are equal to the $\rho$ parameters in (1). In that last case, which would perhaps be optimal in various ways, one would have that

$$EF_{t+h|t} = MF_{t+h|t} + \beta X_t\,. \tag{3}$$

This means that the expert takes the model-based forecast as given and simply adds intuition captured by $\beta X_t$. Preferably, such intuition is unpredictable, because if not, the model itself is mis-specified. When the expert also uses past *Y* with weights in (2) different than those in (1), the expert suffers from double counting. This means that he or she neglects that recent exceptional values of *Y* were already discounted in the model-based forecasts, due to the recursive approach.

**The Law of Small Numbers**

What experts need to do while arriving at (2) is a decision-making process that bears similarities with decision-making by gamblers and investors. There is an information set, and the decision-maker has to judge whether there is reason to act. Here an expert sees the model-based forecast, and also has information on variables that might need inclusion in a final forecast. The size of that act again depends on judgement, which for example means setting the value of $\beta$. Like investors and gamblers, the success rate of the decision is quickly noticeable. Forecasts can be matched with realizations, and a horserace between models and experts can be run.

There is much literature on the behaviour of gamblers and financial investors, and one aspect of their behaviour has been recently summarized in Rabin (2002). See also Camerer (1989), De Bondt and Thaler (1987) and Durham, Hertzel and Martin (2005) for detailed empirical results. This aspect is that the behaviour of these actors obeys the Law of Small Numbers [LSN], a regularity first addressed by Tversky and Kahneman (1971). To quote Rabin (2002), the LSN entails that people "exaggerate how likely it is that a small sample resembles the parent population from which it is drawn."

I conjecture that the behaviour of experts who adjust model-based forecasts also obeys the LSN. Experts receive the model-based forecast, and when a recent large *Y* observation occurs, they likely overlook that the effect of this observation is captured by new estimates of $\rho$ in (1) and by new updated forecasts *MF*. So, they likely keep on modifying model-based forecasts for too long. Additionally, specific information included in *X* is likely to be interpreted as being relevant for future forecasts, more often and longer than necessary. Indeed, shocks are typically absorbed in new values of *Y*, as all actors included such events in their expectations, which in turn leads to modified values of *Y.* Note that this is the basic notion of the Lucas critique. In sum, experts who adjust model-based forecasts are prone to see changes in levels and in trends, based on recent events, which are not there for that long.

## 2. Empirical methodology

Even though there is a large literature on the LSN for gamblers and investors, there is not an overwhelming amount of empirical documentation. Perhaps this is due to the fact that LSN-like behaviour is difficult to observe from actual behaviour. For the case of experts who adjust model-based forecasts, I claim that a rather simple methodology can be designed to elicit this behaviour.

The line of thought is as follows. When an expert incorrectly believes there is a changing trend or level, he or she will consequently incorrectly predict events at some horizon *h*, and it does not matter if that horizon comes closer or not. So, if one combines forecasts for an event at time *t+h*, based on forecasts made at time *t*, *t+1*, *t+2* and so on, the expert who obeys the LSN would systematically do worse than any regression-based model. Indeed, such a regression-based model would each time have a tendency to predict towards the mean of past observations, and systematic under- or over-prediction for the same horizon shall not occur.

To highlight the LSN for experts, I consider forecasts from models and experts when combined for the *h*-step-ahead horizon. This can be written as

$$CMF_{t+h} = \sum_{j=0}^{h-1} \lambda_j MF_{t+h|t+j} \tag{4}$$

which is the weighted sum of model-based forecasts for one and the same horizon h, and

$$CEF_{t+h} = \sum_{j=0}^{h-1} \lambda_j EF_{t+h|t+j} \tag{5}$$

which similarly is the weighted sum of expert forecasts. The weights $\lambda_j$ indicate the relative importance of the different forecasts.

The conjecture in this paper is that if experts' behaviour obeys the Law of Small Numbers that then the combined forecast *CEF* in (5) for horizon *h* is worse than the one-step-ahead forecast made at *h*-1, while for the combined model-based forecast in (4) the success rate should be 50-50.

## 3. Evidence

To examine the validity of the LSN conjecture, I analyse the behaviour of thirty-seven experts located in an equal amount of countries. The experts are the local managers in these countries, who receive forecasts from the headquarters' office and are allowed to modify these forecasts. The data concern monthly sales of pharmaceutical products in seven distinct product categories and cover twenty-five months in 2004 to 2006. Sometimes these categories for some countries concern just a single product, sometimes there are thirty products. In total there are 207 country-category combinations that can be used for further analysis. A closer look at the personal characteristics of the experts indicates that they are mostly male, although there are quite a number of females, that their age distribution is reasonably uniform, as is the distribution of the years of experience.

### Testing for differences across forecast accuracy

The focus is on comparing the forecast accuracy of (4) and (5) with the simple one-step-ahead forecast for $t+h$ made from $t+h$-1, that is $MF_{t+h|t+h-1}$. To test the null hypothesis that the mean squared prediction error of (4) and (5) is equal to that of the one-step-ahead model-based forecast against the alternative hypothesis that (4) or (5) is better, I need to take account of the fact that (4) and (5) nest $MF_{t+h|t+h-1}$ . As is explained in Clark and McCracken (2001), due to this nesting property the relevant test statistic does not have a standard normal distribution. Therefore, the following procedure is followed. I have $P$ recursively created out-of-sample forecasts, with P = 25, while the first forecast is generated using about 60 in-sample data. Hence, the fraction of forecasts over in-sample data is $\pi = P/R = 0.4$, approximately. Next, it is unknown how many variables are included in the additional set of regressors, but for convenience I set $k$ at 2. Based on simulations concerning empirical size and power, Clark and McCracken (2001) recommend using the so-called ENC-NEW test, defined by

$$ENC-NEW = P\frac{\frac{1}{P}\sum(u_{1,t+1}^{2}-u_{1,t+1}u_{2,t+1})}{\frac{1}{P}\sum u_{2,t+1}^{2}} \qquad (6)$$

The summation runs for the $P$ forecasts, and $u_{1,t+1}$ denotes the forecast errors for $MF_{t+h|t+h-1}$ and $u_{2,t+1}$ either (4) or (5). The 5% critical values of the one-sided tests are given in Table 1 of Clark-McCracken (2001, page 92). For $\pi$ = 0.4 and $k$ = 2 it is 1.481. The test statistic in (6) is computed for all products in all country-category combinations. When there is more than one product, the test values are averaged, and the final average value is compared with the critical value.

**Evaluating combined forecasts**

There are 207 country-category combinations, and I will count the number of significantly better forecasts made by (4) or (5), over just $MF_{t+h|t+h-1}$. When using a binomial test, the 95% confidence interval is then 89 to 118, or in terms of fractions 0.43 and 0.57.

The managers at the headquarters' office informed me that the $h$ = 6 horizon is an important horizon for reasons of supply chain management. Also, the bonus payments of the local managers depend on their forecasts, predominantly for this horizon. The first set of results concerns the case where the weights $\lambda_j$ are all equal to 1/6. Equal weights are often found to be best when combining forecasts, see Timmermann (2006). For the combined model-based forecasts in (4) the fraction of cases where (4) has significantly better MSPE is 0.53. In contrast, for the combined expert forecasts this fraction is only 0.41. Hence, the six averaged forecasts for the same horizon are significantly worse than a one-step-ahead forecast based on the previous month. This shows that experts do not tend to re-direct their forecast back to some average level, whereas of course regression-model-based forecasts do that automatically. In other words, experts keep incorporating events too often for too long a period.

## 4. Conclusion and implications

The main conclusion from this paper is that the behaviour of experts who adjust model-based forecasts, like gamblers and financial investors, obeys the Law of Small Numbers. An additional novelty of this paper is that this conclusion is based on the analysis of a very large and detailed database, which allows for drawing such a generalizing statement.

Two main implications follow from this finding. The first is that experts need to play down their added contribution to the final forecast. Apparently they see trends and changes that are not there, and they include these too often in their final forecast. One way to learn more about this phenomenon is that experts start to write down what it exactly is what they do and why they do it.

A second implication is that there seems room for improving communication about what models effectively can do and what they cannot do. The notion that regression-based models convert forecasts towards the unconditional mean apparently is not well understood. Also, the trap of double counting should thus be avoided.

All in all, it seems that there are opportunities for improving the interaction between models and managers. The data analyzed in this paper deal with managers who are not involved in also constructing the model-based forecasts. Perhaps a little more interaction could be beneficial. On the other hand, if the model-builder and the expert are exactly the same person, there strictly speaking should be no reason to adjust the model- based forecast. Hence, somewhere there should be an optimal level of interaction, and further research is needed here.

**References**

Camerer, Colin F. (1989), Does the basketball market believe in the "hot hand"?, *American Economic Review*, 74, 1257-1261.

Clark, Todd E. and Michael W. Cracken (2001), Tests of equal forecast accuracy and encompassing for nested models, *Journal of Econometrics*, 105, 85-110

De Bondt, Werner F.M. and Richard H. Thaler (1987), Further evidence on investor overreaction and stock market seasonality, *Journal of Finance*, 42, 557-581.

Durham, Gregory R., Michael G. Hertzel, and J. Spencer Martin (2005), The market impact of trends and sequences in performance: New evidence, *Journal of Finance*, 60, 2551-2569.

Rabin, Matthew (2002), Inference by believers in the Law of Small Numbers, *Quarterly Journal of Economics*, 117, 775-816

Timmermann, Allan (2006), Forecast combinations, Chapter 4 in Graham Elliott, Clive W.J. Granger and Allan Timmermann (eds.), *Handbook of Economic Forecasting Volume I*, Amsterdam: Elsevier, pp 135-196.

Tversky, Amos and Daniel Kahneman (1971), Belief in the Law of Small Numbers, *Psychological Bulletin*, 76, 105-110.