Outliers and judgemental adjustment of time series forecasts

Philip Hans Franses

Econometric Institute
Erasmus University Rotterdam

Econometric Institute Report 2008-04

Abstract

This paper links judgemental adjustment of model-based forecasts with the potential presence of exceptional observations in time series. Specific attention is given to current and future additive outliers, as these require most consideration. A brief illustration to a quarterly real GDP series demonstrates various issues. The main focus of the paper is on various testable propositions, which should facilitate the creation and the evaluation of judgemental adjustment of time series forecasts.

February 27 2008

Address for correspondence is Econometric Institute, Erasmus University Rotterdam, PO Box 1738, NL-3000 DR Rotterdam, The Netherlands, franses@few.eur.nl. Thanks are due to Rianne Legerstee and especially Dick van Dijk and Paul Goodwin for helpful comments

1. Introduction and motivation

Consider an expert who has domain knowledge concerning an economic variable *Y* and who needs to make one-step-ahead forecasts for its realizations *y*. This expert is aware of the fact that *y* cannot fully be predicted by its past values, but as a benchmark model he or she relies on an autoregression of order *p*. When considered necessary, the expert can use to judgment to modify the time series model forecast to give a final adjusted forecast. Such an adjustment can be relevant when the expert knows that in the next period there will be a structural level shift that is currently not included in the model. Another reason may be that the expert feels that the most recent observation, which in fact is the forecast origin for the one-step-ahead forecast, is exceptional. This can be due to a possible shift in level that has already started, or to a single-observation outlier. At the time of producing the forecast the expert has no tools to decide what could be the reason for this exceptional data point, but he or she does feel the need to judgementally adjust the model forecast. It is this situation that is considered in this paper. As will be seen, whether an expert uses a simple time series model or a more elaborated econometric model does not matter much for the main results, so for ease of use and discussion I stick to the simpler notation here to illustrate various issues. It also does not matter whether or not the expert is actually involved in model building and parameter estimation.

Suppose the expert has data for the variable *Y* for *n* consecutive periods (weeks, months, quarters), and denote these as y_1, y_2, …, y_n. So, observation y_n is the forecast origin, and suppose the expert wants to forecast y_{n+1}. Suppose further that the expert considers y_n as exceptional. In Section 2 below I will show that when the exceptional observation concerns a so-called innovation outlier [IO] (see Fox, 1975 and Franses, 1998 Chapter 6 for the nomenclature), the expert should *not* adjust the forecast. While when it is a so-called additive outlier [AO] then the adjustment should be a function of the forecast error at time *n.* In Section 3 there will be an illustration to forecasting annual growth in quarterly real GDP for the Netherlands, where the most recent observation (at the time of writing) concerns 2007Q3 is considered as quite exceptional. Section 4 elaborates on the link between outliers and judgemental adjustment and it puts forward a range of testable propositions. Section 5 concludes with practical suggestions which can be applied when one has historical data on adjustment and on model-based forecasts.

2. Adjusting (or NOT) for an exceptional observation at the forecast origin

For ease of notation I consider an autoregression of order 1 [AR(1)], that is

(1) $$y_{n+1} = \mu + \rho y_n + \varepsilon_{n+1} .$$

The forecast made at origin n is equal to

(2) $$\hat{y}_{n+1|n} = \mu + \rho y_n$$

It is assumed that the parameters can be consistently estimated using ordinary least squares [OLS]. Ledolter (1989) has shown that if an exceptional observation (whether it is an IO or an AO that does not matter) occurs at the forecast origin this does not have much effect on the OLS estimated parameters. To save notation, I therefore use hats only for forecasts, and not for estimated parameters.

One example of an exceptional observation is an innovation outlier [IO], which when it occurs at the time n can be written as

(3) $$y_n = \mu + \rho y_{n-1} + \varepsilon_n + \omega_n$$

where ω_n is defined as equal to ω at time n and equal to 0 elsewhere. Assuming that $\omega > 0$, this means that y_n is exceptionally larger at time n even when taking the distribution of ε_n into account. Now, the interesting feature of an innovation outlier is that for the next observation one simply has that

(4) $$y_{n+1} = \mu + \rho y_n + \varepsilon_{n+1}$$

and so, when looking back from n+1, there is *nothing exceptional* about y_n when forecasting y_{n+1}, while y_n *is* exceptional when seen from origin n-1. Hence, in case of an innovation outlier, the expert better *not* adjust the forecast, even though the value of y_n is being thought of as exceptional. Of course, at time n it is unknown whether the observation is an IO or not.

The discussion changes dramatically for the case of a so-called additive outlier [AO] at time n. For the AR(1) model this would mean that the true data are

(5) $$y_n = \mu + \rho y_{n-1} + \varepsilon_n$$

but that one does not observe realizations of Y but of another (but of course very related) variable X, with realizations

(6) $$x_n = y_n + \omega_n$$

with ω_n again defined as equal to ω at time n and equal to 0 elsewhere, suppose. Note that this means that the observed data are $\{y_{n+1}, x_n, y_{n-1}, y_{n-2}, \ldots\}$. Substituting the AR(1) expression (5) gives for time n

(7) $$x_n = \mu + \rho y_{n-1} + \varepsilon_n + \omega_n$$

which means that the forecast error at time n is $\varepsilon_n + \omega$, and for time n+1

(8) $$x_{n+1} = \mu + \rho x_n + \varepsilon_{n+1} = \mu + \rho(y_n + \omega) + \varepsilon_{n+1} = \mu + \rho y_n + \varepsilon_{n+1} + \rho\omega$$

Thus, for x_{n+1} to become y_{n+1}, as it should, the expert would provide the most successful adjustment if he or she would *subtract* $\rho\omega$ from the model-based forecast made at time n.

In case of an AO, and as the expected value of ε_{n+1} is zero, the expert-adjusted forecast reads as

(9) $$\hat{y}^a_{n+1|n} = \hat{x}_{n+1} - \rho\omega$$

The question of course is what the value of ω is. One choice could be to set it equal to the forecast error $e_{n+1|n}$, that is, the error observed at time n, when the forecast is made at time n-1. In that case,

(10) $$\hat{y}^a_{n+1|n} = \hat{x}_{n+1} - \rho(\varepsilon_n + \omega) = \mu + \rho y_n - \rho(y_n - \rho y_{n-1} - \mu) = (1+\rho)\mu + \rho^2 y_{n-1} = \hat{y}_{n+1|n-1}$$

So, assuming that the observation itself at time *n* can be viewed as an AO effectively means that the expert-adjusted forecast is the two-step-ahead forecast from origin *n*-1.

Perhaps a more plausible option to calibrate ω (when $\omega > 0$) is by

(11) $$\hat{\omega} = e_{n|n-1} - k\sigma$$

where σ is the standard error of ε_n and k can be set by the expert.

In sum, *only* when the exceptional observation at time *n* is an *additive outlier*, the adjustment made by the expert would result in forecast quality improvement. In fact, when the expert modifies the forecast in case there is an innovation outlier, forecast quality will in fact deteriorate, at least in principle. This intriguing result makes judgemental adjustment a hazardous exercise, as at time *n* one cannot possibly know whether the observation at the forecast origin is an IO or an AO.

3. Illustration

To illustrate the results in Section 2, consider annual growth rates of quarterly real GDP in the Netherlands, for the period 1997Q4 to 2007Q3. The realization in 2007Q3 is 4.2%, and by many this is viewed as exceptionally high. Statistics Netherlands also appreciates this notion by stating that part of this sudden increase in growth rates might be due to increased revenues of natural gas production. The mean value of growth per quarter in the considered period is 2.4.

A suitable model to describe these data is an AR(1) model with parameter ρ estimated equal to 0.866 for the sample ending in 2007Q3 and 0.863 for the sample ending in 2007Q2, supporting the results in Ledolter (1989). The σ is estimated as 0.0072, and the forecast error $e_{n|n-1}$ for 2007Q3 is 0.0154, which means that the actual minus the forecast is 1.54%, which indeed is not small.

The forecast for 2007Q4 based on this AR(1) model would be equal to 3.9%. In that case, the expert adopts the IO notion and does not intervene. Suppose now that it is treated as an AO. If 2007Q3 is fully treated as such, then the adjusted forecast is 3.9-0.866(1.54) = 2.6%. If I would use (11) with $k = 2$, the estimated value for ω is small and the adjusted forecast is 3.9-0.866*(0.1) = 3.8%. With a k equal to 1, the adjusted forecast becomes 3.2%.

In reality the expert who adjusts a forecast should have strong arguments why he or she wants to allow for just one standard deviation.

4. Implications and testable propositions

The discussion and the results in Section 2 may seem rather trivial and obvious. However, in this section it will be argued that there are various consequences for carrying out and evaluating judgemental adjustment to model-based forecasts. A key issue here concerns the very nature of that apparently exceptional observation at the forecast horizon and it is also important how often unexpected data points like these could occur.

Evaluating forecast quality

Now we know how to properly deal with an exceptional observation at the forecast horizon, the next question is of course whether such adjustment would help. This can only be known at time n+1, and hence no sensible statement can be made at the time of making the forecast adjustment. The key reason for this is that one needs to know what kind of exceptional observation y_n was and this can only be learned afterwards.

The only thing that one *can* do is to study past realizations and to examine whether exceptional data, if there are any, typically are innovation or additive outliers. This could give a first impression of what could possibly be happening at time n. Various methods are available for this purpose, and useful examples are the methods outlined in Chen and Liu (1993) and Tsay (1988).

Proposition 1:

If only once in a while an *innovation outlier occurs at the forecast origin*, there is no need to adjust the model-based forecast. In fact, adjustment would lead to poor expert forecasts.

Proposition 2:

If only once in a while an *additive outlier occurs at the forecast origin*, judgemental adjustment by subtracting or adding a fraction of the recent forecast error should lead to substantial forecast quality improvement.

These two propositions are testable using actually observed time series data, the model-based forecasts and the expert-adjusted forecasts. Of course, also simulations can be used to show the empirical relevance of these two propositions.

Several innovation outliers in the future

One possible scenario is that the expert foresees the occurrence of an innovation outlier, simply as these have been observed before and some common tendency seems noticeable. Typically this would be associated with an upcoming change in regulations like taxes, or institutional changes like interest rate changes, or with a combination of unusual factors which do sometimes happen like strong stock market dips or rapid oil price increases. In a sense, the expert may then try to forecast the size of the upcoming innovation outlier, and when this is possible one wants to take account of that. Assuming for the moment that only positive exceptional innovations would occur, a simple adjustment scheme for the expert to follow now would be to consider

$$(12) \qquad y_{n+1} = \mu + \rho y_n + \varepsilon_{n+1} + \omega_{n+1}$$

with

$$(13) \qquad \omega_{n+1} = \alpha + \beta z_n + \eta_{n+1},\ \textit{if}\ \omega_{n+1} > 0,\ \textit{and}\ \omega_{n+1} = 0,\ \textit{otherwise}$$

where z_n covers explanatory variables and to use as a forecast

$$(14) \qquad \hat{y}_{n+1} = \mu + \rho y_n + \hat{\omega}_{n+1}$$

The expression in (13) is a so-called censored regression model for the outlier component, and this combined model has been introduced in Franses and Paap (2002) as the censored latent effects autoregression. The expert needs to decide on the values of α and β and on the variance of η, where of course the choice for the variables in z is also important. When positive and negative innovation outliers can occur, one can replace the expression in (13) by an expression with two-sided censoring. A useful by-product of this way of formulating what an expert can do is that it helps to keep track of what experts actually do when they modify model forecasts. A plea for keeping track of their behaviour has been made in a variety of recent studies on judgemental adjustment to forecasts, see Sanders and Ritzman (2001) and Lawrence et al (2006), to mention just a few. Of course, one may also want to modify the actual time series model and replace it by (12) with (13) in case innovation outliers occur on a regular basis. Looking at the expression of an IO in (3), the occurrence of frequent innovation outliers can be noted from as many large but isolated forecast errors, see also Franses (1998, Chapter 6).

Proposition 3:
If innovation outliers occur quite frequently, and they are to some extent predictable, one can formulate judgemental adjustment in a censored latent regression framework.

Several additive outliers

In case one expects the frequent appearance of additive outliers, one basically has the situation where part of the observations obeys another data generating process [DGP]. To illustrate, consider the following DGP for a quarterly observed variable

(15) $$y_n = \alpha + \alpha_1 D_{1,n} + \varepsilon_n$$

where $D_{1,n}$ is a dummy variable which takes the value 1 in quarter 1 and 0 elsewhere. So, the mean of the variable is α in quarters 2, 3 and 4, but it is $\alpha + \alpha_1$ in quarter 1. Another way of putting it is to say that y_n has an error term with mean 0 and variance σ^2 in quarters 2, 3 and 4 and an error term with mean α_1 and variance σ^2 in quarter 1. A graph of this series would indicate seasonality, which can also be viewed as a repetitive sequence of additive outliers.

What would happen if one used an autoregression to describe this y_n? Given the quarterly seasonality, it could be a model with y_{n-4} as the explanatory variable. The parameter to be estimated is then ρ_4. For most of the observations the true value of this ρ_4 is of course 0, and in fact the whole model is mis-specified. The estimated forecast errors for this model will therefore (on average) be positive for quarter 1 and negative for quarters 2, 3, and 4. The empirical distribution of the forecast errors therefore shows fatter tails and is also skewed. Looking at these errors and knowing that the model misses out on regular seasonality the expert observes that the model is mis-specified as there are so many additive outliers, and hence he or she will almost always adjust and in fact according to a rather systematic pattern.

Proposition 4:

If it is observed that an expert typically exercises one-sided adjustment (for example, most often upwards), then this could be in accordance with asymmetric past model-based forecast errors.

Proposition 5:

If it is observed that an expert adjusts almost all model-based forecasts, then, given the time series model used, the data should show many additive outliers.

Note that these two propositions assume that the expert is not just randomly adjusting and also that he or she sometimes does not adjust a model-based forecast. Proposition 5 basically says that if there are so many additive outliers, then the model *mus*t be mis-specified, and this shall be known to the expert. Note that this does not mean that the expert can claim that *because* the model is a simple time series model it *must* be wrong, as it could well be that just occasional innovation outliers are at stake and hence that on average the model is not bad at all. If the time series model is not useful on its own, that it should be detected from the presence of many additive outliers, and only then, the expert can claim that he or she often or almost always has to adjust.

Model mis-specification

When there are too many additive outliers, almost everything goes wrong for the time series model, see Ledolter (1989), Franses (1998, Chapter 6). Parameters are estimated incorrectly, forecast errors are very large and their distribution is likely to be skewed, as indeed the forecasts are biased. Of course, the key reason for this is that so many additive outliers basically imply that the model is not well specified. For example, consider the case

(16) $$x_n = y_n + \omega_n$$

and where the *Y* variable obeys an AR(1), and where the additive outliers can be described by

(17) $$\omega_n = \alpha + \beta z_n + \eta_n$$

Writing y_n as $x_n - \omega_n$, one then has

(18) $$x_n - \alpha - \beta z_n - \eta_n = \mu + \rho(x_{n-1} - \alpha - \beta z_{n-1} - \eta_{n-1}) + \varepsilon_n$$

or, equivalently,

(19) $$x_n = \mu + (1-\rho)\alpha + \rho x_{n-1} + \beta z_n - \rho\beta z_{n-1} + \varepsilon_n + \eta_n - \rho\eta_{n-1}$$

which is a so-called ARMAX(1,1,1) model, as it includes lagged explanatory variables and it has a first-order autocorrelated error term (a moving average, MA). Interestingly, this model with an MA component is quite frequently found for macroeconomic time series variables when there are measurement errors (which is about the same as additive outliers), see Granger and Morris (1976).

So, when additive outliers occur very frequently, this is usually a sign of serious misspecification of the model. It still is possible for an expert to add value based on domain knowledge, but (19) also indicates that for this added value to lead to better forecasts one needs to specify a rather complicated expression based on lagged variables and (perhaps unobserved) error terms. In fact, I would conjecture that it would seem quite unlikely that the expert has *that* much knowledge of the process that he or she is able to construct (19).

Proposition 6:
When there are too many additive outliers, the quality of the adjustment by the expert most likely decreases as the optimal forecasting scheme is very complicated.

To sum up, when a simple time series model is used by an expert and occasional innovation outliers occur, the expert may be most successful if he or she somehow is able to forecast their value, based on domain-specific information. In contrast, when often additive outliers occur, it seems best to re-specify the model, for example by entering additional variables.

Time series properties of the expert's contribution

Given the above results for additive and innovation outliers, it seems that experts' added value to model-based forecasts would be most beneficial to forecast accuracy if experts only once in a while have to intervene. If adjustment is always felt needed, then the model clearly is felt to lack important variables. So often providing adjustment simply cannot be always beneficial, and the simple advice should be that the model needs to be re-created. Interventions that could be most successful would respond to an additive outlier at the forecast origin, or to a future innovation outlier, foreseen by the expert. All in all, the frequency of expert adjustment at best is low. Note that the size and the sign of adjustment do not matter much, because this depends on the nature of the outliers that can occur. So, observing more positive adjustments by an expert is not necessarily a bad sign. Also, because adjustment happens only once in a while, at least in the optimal situation, it is quite unlikely that expert's adjustment can be predicted.

To reverse this argument, it can be seen as *not* a good sign in case the differences between the model-based forecasts and the final expert forecasts are predictable from explanatory variables as that would mean that the time series model is considered by the expert as seriously mis-specified. Given the argument in Section 2, one would also not want that the experts' added values are serially correlated, as that would mean that the experts treat the additive outlier mechanism as an autoregression itself. For example, if the expert knows that the model forecasts are obtained from

(20) $$y_n = \mu + \rho_1 y_{n-1}$$

but the expert also knows that this model is not well specified because it lacks the explanatory variable x, he or she might use instead

(21) $$y_n = \alpha + \lambda_1 y_{n-1} + \beta x_n$$

as a forecast, for which it is quite unlikely that $\rho_1 = \lambda_1$. In that case, the added contribution of the expert is

(22) $$\mu - \alpha + (\lambda_1 - \rho_1) y_{n-1} + \beta x_n$$

This added value could be more often positive than negative, and it also will show correlation over time. The question here is of course whether the expert is able to properly calibrate (20). Taking this doubt into account, and also the discussion of Section 2, I would put forward the final testable proposition, that is

Proposition 7:
The more predictable is the adjustment made by the expert, the less likely does the final forecast quality get improved.

Indeed, it seems best to modify the model first, before it is given to the expert for a final touch. And, the less predictable the adjustment by the expert is, the better can be the quality of the final forecast.

5. Conclusion

This paper has shown that the relevance and quality of judgemental adjustment of model-based forecasts depends on the type of outliers in the data. The results were all derived for simple autoregressive time series models, but it is easily understood that they can be extended to more general econometric forecasting models. The key results are that innovation outliers at the forecast horizon do not require judgemental adjustment, but that, if is possible, future such outliers can be included in an adjustment scheme. In case of additive outliers at the forecast horizon, judgemental adjustment should improve forecast quality, but when there are too many of these, it is better to re-specify the model as it is quite unlikely that experts can come up with the more appropriate but rather complicated forecasting scheme.

The practical consequence of this study is that if one wants to evaluate the quality of judgemental adjustment of model-based forecasts it helps to first study the time series properties of the data, and to study the occurrence of outliers. A second issue for future work amounts to a re-iteration of earlier recommendations and that is that it is very useful to document exactly what the expert does. This is relevant for the expert, but also for the modeller, who might learn in what way the model can be improved. Franses and Legerstee (2007) have provided some empirical evidence on what experts actually do, but more evidence is needed.

References

Chen, C. and L.-M. Liu (1993), Joint estimation of model parameters and outlier effects in time series, *Journal of the American Statistical Association*, 88, 284-297.

Fox , A.J. (1972), Outliers in time series, *Journal of the Royal Statistical Society* B, 34, 350-363.

Franses, P.H. (1998), *Time Series Models for Business and Economic Forecasting*, Cambridge UK: Cambridge University Press.

Franses, P.H. and R. Legerstee (2007), Dynamics of expert adjustment to model-based forecasts, Revision requested by the *International Journal of Forecasting.*

Franses, P.H. and R. Paap (2002), Censored latent effects autoregression, with an application to US unemployment, *Journal of Applied Econometrics* 17, 347-366

Granger, C.W.J. and M.J. Morris (1976), Time series modelling and interpretation, *Journal of the Royal Statistical Society* A, 139, 246-257.

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

Ledolter, J. (1989), The effect of additive outliers on the forecasts from ARIMA models, *International Journal of Forecasting*, 5, 231-240.

Sanders, N. and L. Ritzman (2001), Judgemental adjustments of statistical forecasts, in J.S. Armstrong (ed.), *Principles of Forecasting*, New York: Kluwer

Tsay, R.S. (1988), Outliers, level shifts, and variance changes in time series, *Journal of Forecasting*, 7, 1-20.