Discussion on
"Probability Distributions or Point Predictions? Survey Forecasts of US Output Growth and Inflation"
by Michael Clements

Discussant: Kajal Lahiri
University at Albany:SUNY

Workshop on Uncertainty and Forecasting in Macroeconomics,
June 1-2, 2012. Eltville

Engelberg, Manski and Williams (JBES 2009) first documented that point forecasts and means from density forecasts are sometimes systematically different – point forecasts seem to have systematic favorable bias. Based on this observation, they suggested emphasizing density forecasts more than point forecasts.
Clements (EER 2010) found that point forecasts are more accurate than means of density forecasts in terms of MSE criterion. Current paper is a continuation of that theme using a different metric.

First, Clements documents that means derived from density forecasts do not follow simple consistency over horizons that is dictated by a simple Bayesian Learning Model (BLM), whereas point forecasts do much better on that account. As forecasts get closer to the target, more importance should be placed on new information compared to initial prior beliefs.

Second, he examines if different ways of approximating histograms, e.g., Uniform, Normal, Beta, etc. make much difference in establishing the result that density forecasts are not well calibrated. By improving the methodology of PIT evaluation, he establishes that these densities are not well calibrated. Hence, the point he makes is that we should continue to ask for and use point forecasts.

His basic point is well taken.

Whereas I am impressed by the second part of the paper, and find the results expected and believable, I will restrict my main comment to the first and major part of the paper: I have some concerns on the model formulation and estimation of the BLM. I have two major points.

If the purpose is the examine if certain consistency and optimality properties are satisfied in a sequence of fixed target forecasts, there are simple tests that one can implement based on results from Lahiri and Isikler (IJF 2006), Patton and Timmermann (JBES 2012), Lahiri (JBES 2012), etc,. For instance, variance of forecast errors should fall and variance of forecasts should increase as the forecast horizon decreases. One can also do Nordhaus type tests horizon by horizon - Clements has done much early work of the latter test. These simple yet powerful tests can be done concurrently using both point and means of density forecasts. More generally, using the whole densities, one can test for significant information gains over successive horizons using Kulback-Leibler and other entropy measures.

The Bayesian Learning model that Clements is working with is not consistent with point forecasts either – see Lahiri and Sheng (2008, 2010). Then one can possibly show that point forecasts are more in line with BLM than the means of density forecasts are.

Let me explain why I do not think that BLM has not been implemented properly.

Bayesian learning model with Heterogeneous Agents:

Assumption on prior beliefs:

The initial prior belief of the target variable for the year *t*, held by the forecaster *i*, at the 24-month horizon is represented by the normal density with the mean F_{it24} and the precision (i.e. the reciprocal of the variance) a_{it24} for $i = 1,\ldots,N,\;\; t = 1,\ldots,T.$

Assumption on public information arrival:

At horizon *h* months, experts receive one public signal L_{th}, but do not interpret it identically. In particular, individual *i*'s estimate Y_{ith}, conditional only on the new public signal, can be written as

$$Y_{ith} = L_{th} - \varepsilon_{ith}, \quad \varepsilon_{ith} \sim N(\mu_{ith}, b_{ith}).$$

ε_{ith} - expert *i*'s error term

μ_{ith} - expert *i*'s interpretation of the public information

b_{ith} - perceived quality of public information by expert *i*

Bayes rule implies that under the normality assumption, agent *i*'s posterior mean is the weighted average of his prior mean and his estimate of the target variable conditional only on new public signal:

$$F_{ith} = \lambda_{ith} F_{ith+1} + (1 - \lambda_{ith})(L_{th} - \mu_{ith}),$$

with his posterior precision $a_{ith} = a_{ith+1} + b_{ith}$, where $\lambda_{ith} = a_{ith+1} / (a_{ith+1} + b_{ith})$ is the weight attached to prior beliefs.

For convenience, the following population parameters are defined across professional forecasters for target year *t* at horizon *h*:

$E_i(F_{ith}) = F_{th}$, $\mathrm{var}_i(F_{ith}) = \sigma^2_{F|th}$: differences in their prior beliefs, $\sigma^2_{F|th+1}$;

$E_i(\lambda_{ith}) = \lambda_{th}$, $\text{var}_i(\lambda_{ith}) = \sigma^2_{\lambda|th}$: differences in the weights attached on priors,

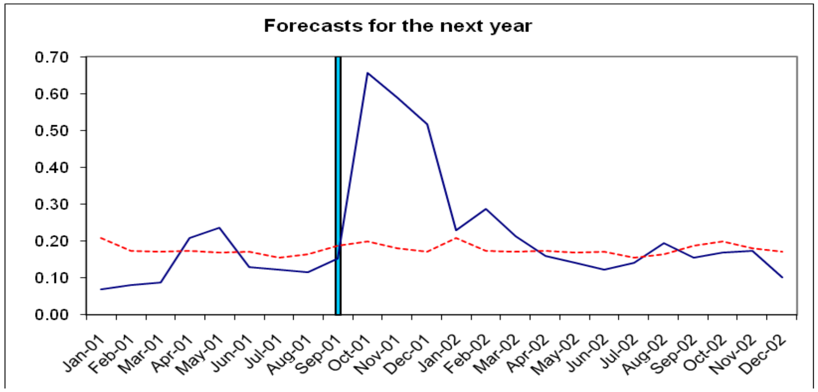
$E_i(\mu_{ith}) = \mu_{th}$, $\text{var}_i(\mu_{ith}) = \sigma^2_{\mu|th}$:differential interpretation of public information, $\sigma^2_{\mu|th}$.

where $\lambda_{ith} = a_{ith+1}/(a_{ith+1} + b_{ith})$ is the weight attached on prior beliefs.

Clements, following Manzan (2011), uses the last known value as a proxy for L_{th}. This is typical in the Accounting literature. Given the very definition of L_{th}, the approach is doomed to fail.

One can never be sure that the last announced actual value is common information to all, and that there was no differential information between forecasters. In addition, many high frequency intra-quarterly indicators are used for forecasting. So the parameter can not be interpreted as differential interpretation of common public information.

Differential interpretation of public information
– Effect of 9/11 on the disagreement in US GDP forecasts



Essentially one has a dynamic panel data model with heterogeneity. Pesaran's (Econometrica 2006) Common Correlated Effects Estimator was not meant for dynamic models. So I worry about the consistency property of the estimator adopted by Clements.

Estimation of $(1-\lambda_h)$, the weight attached to public information:

Lahiri and Sheng (JE 2008, IJF 2010) outline a different way of estimating λ.

$F_{ith} = \lambda_{ith} F_{ith+1} + \varepsilon_{ith}$, where $\varepsilon_{ith} = (1-\lambda_{ith})(L_{th} - \mu_{ith})$. By construction, ε_{ith} and F_{ith+1} are independent for any *t* and *h*.

First, note that above is not estimable, since the number of parameters to be estimated exceeds the number of observations. We assume

$$\lambda_{ith} = \lambda_{ih} = \lambda_h + v_{ih},$$

where v_{ih} has mean zero, mutually independent of each other, and independent over forecast horizons. We regress the forecast revision (ΔF_{ith}) on the lagged forecast (F_{ith+1}) to circumvent the possible problem of spurious regression.

Thus, the estimable version becomes

$$\Delta F_{ith} = \beta_h F_{ith+1} + u_{ith},$$

where $\beta_h = \lambda_h - 1$ and $u_{ith} = \varepsilon_{ith} + v_{ih} F_{ith+1}$.

Second, u_{ith} will be correlated across forecasters because, conditional on *t* and *h*, it has a nonzero mean that depends on L_{th}. To solve this problem, we rewrite the equation as

$$\varepsilon_{ith} = \Delta F_{ith} - \beta_h F_{ith+1} - v_{ih} F_{ith+1} .$$

Taking expectations over *i* conditional on *t* and *h*, we get

$$E_i(\varepsilon_{ith}) = \Delta F_{th} - \beta_h F_{th+1} .$$

Subtracting, we obtain

$$\Delta F_{ith} - \Delta F_{th} = \beta_h (F_{ith+1} - F_{th+1}) + w_{ith} ,$$

where $w_{ith} = \varepsilon_{ith} - E_i(\varepsilon_{ith}) + v_{ih} F_{ith+1}$. In contrast to u_{ith} , the error w_{ith} has a zero mean now.

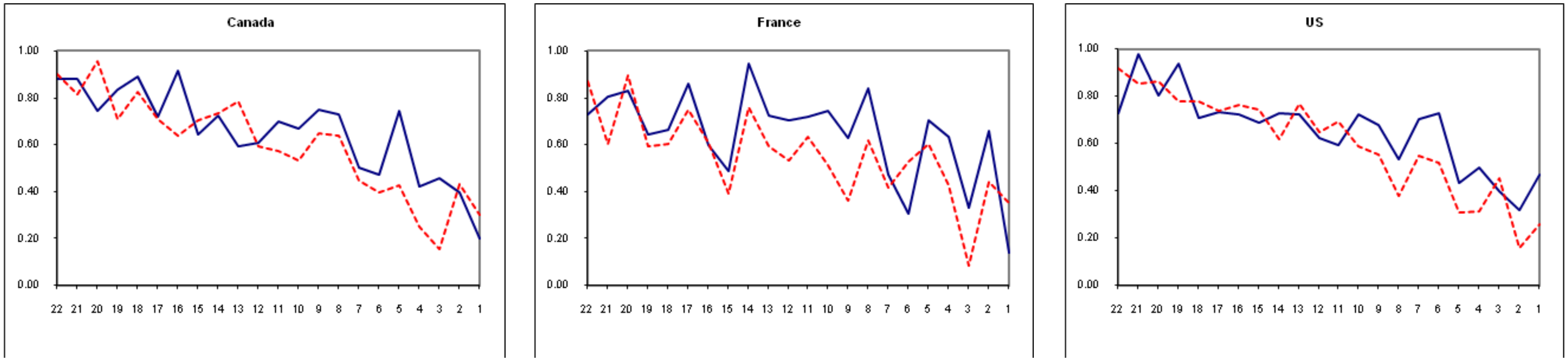
Finally, w_{ith} might be serially correlated. Let w_{ith} follow the AR(1) process $w_{ith} = \rho_h w_{ith+1} + \eta_{ith}$, then it can be rewritten as

$$\Delta F_{ith} - \Delta F_{th} = \beta_h (F_{ith+1} - F_{th+1}) + \rho_h (\Delta F_{ith+1} - \Delta F_{th+1}) - \rho_h \beta_h (F_{ith+2} - F_{th+2}) + \eta_{ith} ,$$

where $E(\eta_{ith}\eta_{i't'h'}) = \sigma^2_{\eta(i)}$, for $i = i', t = t', h = h'$ and 0 otherwise. Using nonlinear least squares, we estimate the above equation for each horizon after controlling for the heterogeneity in the error term.

Clements has too many parameters with small number of observation for each forecaster. Joint GLS estimation requiring inversion of a big Ω is an over kill and will invite specification errors. Following Lahiri and Sheng (IJF 2010), λ can be estimated for each forecaster for each horizon using point and means from density forecasts, and can be tested for equality. This approach is simpler and avoids the use of proxy for common public information L_{th}.

Contribution of initial prior beliefs in explaining GDP (solid line) and inflation (dotted line) disagreement (Lahiri and Sheng IJF 2010).



Note from above that the parameter value does not fall uniformly over horizons, it depends also when important information arrives.

Asymmetric Loss:

One major way that point forecasts may seemingly look worse that means from density forecasts in terms of MSE criterion if the loss functions are of forecasters are other than simple MSE. Papers by Granger, Diebold, Elliott, Komunjer and Timmermann (EKT) and others have demonstrated that optimal point forecasts can be different from measures of central tendencies of density forecasts. Each professional forecaster may have many clients whose loss functions can be different. The professional forecaster generates a true forecast density, and then delivers optimal point forecasts depending on loss functions.

First, Following Engelberg et al., a bounds analysis should be conducted.

Point forecasts vs., the central tendency of density forecasts

Real output growth (1981Q3-2010Q4)				
	4Q Ahead Forecast	**3Q Ahead Forecast**	**2Q Ahead Forecast**	**1Q Ahead Forecast**
Mean	421/0.0356/0.8195/0.1449	495/0.0364/0.8364/0.1273	518/0.0405/0.8745/0.0849	570/0.0246/0.9439/0.0316
Median	557/0.0952/0.7325/0.1724	660/0.0682/0.7788/0.1530	616/0.0860/0.8247/0.0893	646/0.0759/0.8746/0.0495
Mode	555/0.0631/0.8468/0.0901	656/0.0442/0.8735/0.0823	614/0.0537/0.8893/0.0570	645/0.0434/0.9178/0.0388
Inflation (1968Q4-2003Q4)				
	4Q Ahead Forecast	**3Q Ahead Forecast**	**2Q Ahead Forecast**	**1Q Ahead Forecast**
Mean	930/0.2151/0.7118/0.0731	992/0.1996/0.7550/0.0454	861/0.1359/0.7677/0.0964	627/0.0973/0.8756/0.0271
Median	1110/0.2468/0.6649/0.0883	1200/0.1958/0.7342/0.0700	1008/0.1647/0.7004/0.1349	717/0.1060/0.8257/0.0683
Mode	1108/0.1606/0.7735/0.0659	1195/0.1339/0.8192/0.0469	1005/0.1154/0.7861/0.0985	714/0.0784/0.8838/0.0378

Stylized facts about loss functions

- Most point forecasts fall within the bounds on the central tendency (mean, median and mode). But still for a significant fraction of observations, they do not. This fraction is usually between 5% and 25% and varies over forecast horizons and across different measures of the central tendency.

- Forecasters who skew their point forecasts tend to present rosy scenarios. For real output growth, forecasters are more likely to report a point forecast that is above the upper bound on the central tendency; for inflation, however, forecasters are more likely to report a point forecast that is below the lower bound on the central tendency.

- **As the forecast horizon shortens, the point forecasts are more consistent with the central tendency of the density forecasts. This empirical regularity has not been addressed by Clements.**

- Above facts imply that for inflation, underprediction is less costly than overprediction while for real output growth, underprediction is more costly than overprediction. In addition, the longer the forecast horizon, the more asymmetric is the loss function.

Clements (2011) has argued that the deviation of the point forecast from the central tendency of the density forecast is not due to asymmetric loss function because the point forecast is more accurate than the mean of the density forecast in terms of lower mean squared error. Note that under asymmetric loss, the means of the density forecasts need not necessarily to be more accurate than the corresponding point forecast in terms of MSE. The mean of the density forecasts minimizes the mean squared error with respect to the subjective density, not the objective density. Suppose the density forecast is f(y), the true data generating process is p(y), the mean of f(y) is m, the corresponding point forecast is y^*, and the mean squared error loss function is $(y^h-y)^2$. Then m minimizes $\int(y^h-y)^2*f(y)dy$, not $\int(y^h-y)^2*p(y)dy$. Therefore the ex post mean squared error of m need not necessarily to be smaller than that of the optimal point forecast, y^*. For example, suppose the true data generating process is N (3, 1), and the density forecast is N(1,1). If the loss function is asymmetric such that the cost of underprediction is higher than that of overprediction, the optimal point forecast will be larger than the mean of the density forecast, say 2. In this case the point forecast is closer to the mean of the true data generating process than the mean of the density forecast. This implies that the mean squared error of the point forecast will be lower than that of the mean of density forecast although the point forecast is optimal based on forecaster's loss function and density forecast.

In my recent work with Fushang Liu, we are looking closely into asymmetries in loss functions by using both individual density forecasts and point forecasts without assuming rationality and time invariance as in Elliott et al.

"On the estimation of Forecasters Loss Function Using Density Forecasts," Proceedings of the Business and Economic Statistics Section, American Statistical Association Annual Meetings, 2009.

I feel we can not yet rule out the role of asymmetry in generating point forecasts derived from underlying density forecasts. In a world with many forecasters each having multiple clients with diverse objective functions, this seems reasonable. In the psychology literature, evidence in abundant to this effect.