Chapter 1 Discussion

In this chapter, we discuss the scope and application of the proposed human-machine ensemble method.

1.1 Scope of Human-Machine Ensemble

The experiment applied the proposed human-machine ensemble method to actual economic forecasts, especially inflation forecasts. As a result, in the majority cases, the ensemble method forecasted more accurately than a time-series model, a machine alone and a survey alone. However, it did not always make accurate forecasts, sometimes made worse forecasts than a survey. Based on this result, this section discusses the conditions for applying the human-machine ensemble method.

First, the forecast target is a numerical value. In the human-machine ensemble method, since the ensemble combines humans and a machine so as to minimize the expected error, it is necessary to forecast numerical values that can be defined errors. However, if the error can be defined, it is possible to extent the proposed method to other problems than regression such as classification. Even in that case, the ensemble should use the machine if the expected error of the machine is small, and use humans if it is large.

Second, the prediction model used as a machine model in the ensemble method must be able to output the expected error according to input. Since an ARMA model assumes the process is stationary, the variance of forecasts is constant regardless of the input. Therefore it can not be used as a machine in the ensemble method.

Third, it is necessary to estimate parameters $\mathrm{var}(\varepsilon_h)$ and $\mathrm{cov}(\varepsilon_h)$ of human population in advance. In the experiment, we estimated these parameter values from long-term survey data, the Livingston Survey and the SPF. The parameter estimation is also possible when a firm makes demand forecasts by known people. Even when using humans whose parameters are unknown such as crowdsourcing, for example, an expectation-maximization (EM) algorithm performs parameter estimation. However, it is difficult to adopt this method

to forecasts where the true value is obtained only quarterly or monthly, such as economic forecasts.

Finally, the most important factor is the relationship between humans and a machine. It is necessary for the ensemble method that a machine is more accurate in some cases while a group of humans is more accurate in other cases. If the machine always makes more accurate forecasts than the group of humans, you should use the machine only; if the group of humans always makes more accurate forecasts than the machine, you should not use the machine. Also, if errors of the machine and humans are always similar, the proposed method does not perform well. This is because if the error of humans is also large when the expected error of the machine is large, it is meaningless to combine them. However, if there are cases where either the machine or group of humans is accurate, the proposed method performs well.

Additionally, from the theoretical and empirical results, $\mathrm{cov}(\varepsilon_\theta, \varepsilon_h)$ should be sufficiently less than $\mathrm{cov}(\varepsilon_h)$. It is necessary to satisfy equation (**??**) for the ensemble to adopt the number of humans between zero and maximum. That is, if the diversity of the humans and machine is sufficiently large relative to diversity of only humans, the ensemble can adopt various number of humans. However, the all seven ensembles used in the experiment did not satisfy equation (**??**), and the three of them, in which $\mathrm{cov}(\varepsilon_h)$ was larger than $\mathrm{cov}(\varepsilon_\theta, \varepsilon_h)$, did not adopt any machine forecasts at all, or made worse forecasts than those of only humans. Thus, it is not necessary to satisfy equation (**??**), but it is necessary that $\mathrm{cov}(\varepsilon_\theta, \varepsilon_h)$ is less than $\mathrm{cov}(\varepsilon_h)$ to use the human-machine ensemble method efficiently.

1.2   Application of Human-Machine Ensemble

This section describes the human-machine ensemble method with crowdsourcing and the merits other than forecast accuracy of the human-machine ensemble method.

Crowdsourcing is available as humans in the proposed method instead of the Livingston Survey and the SPF. Crowdsourcing is derived from *crowd* and *outsourcing*, and means outsourcing tasks to an undefined large group of

people [**?**]. Platforms for crowdsourcing such as Amazon Mechanical Turk are on the Web, and you can easily use the labor force on them. The advantage of using crowdsourcing for the human-machine ensemble method is that you can know the required number of workers in advance. On the other hand, whereas the surveys such as the Livingston Survey and the SPF employ specific experts, the crowdsourcing employs unspecific ordinary people. However, Michigan survey, which is a survey for economic forecasts, employs ordinary people, but makes as accurate forecasts as the Livingston Survey and the SPF. This implies that the similar accuracy can be obtained even if crowdsourcing is used for the ensemble method instead of surveys by experts. It is, of course, necessary to estimate the parameters $\text{var}(\varepsilon_h)$, $\text{cov}(\varepsilon_h)$ and $\text{cov}(\varepsilon_\theta, \varepsilon_h)$. Therefore, it is not possible to apply the ensemble method immediately. You must gather samples for parameter estimation.

The human-machine ensemble method also suggests a form of collaboration for machines and humans. We assumed that the machine makes larger expected error when given a sequence not in the past. Conversely, observing the number of humans that combined in the ensemble method reveals the input that makes the machine error larger. In this way, you can find problems at which the machine is not good but humans are good. It also suggests the direction of improving the machine forecasts, or a new human-machine ensemble method.