

Forecast combination for discrete choice models: predicting FOMC monetary policy decisions

Laurent L. Pauwels^a and Andrey Vasnev^a

^a*The University of Sydney Business School, New South Wales 2006, Australia*
Email: laurent.pauwels@sydney.edu.au

Abstract:

This paper suggests combining multiple forecasts from discrete choice models for the FOMC decisions on the target rate instead of relying on one specific model. The main advantage of combining forecast is to improve forecast accuracy, as first shown by Bates and Granger [1969]. There are several other advantages to forecast combination; for example it allows to deal with a larger number of variables than a univariate model or it may be more robust to potential breaks in trends and intercepts. Timmermann [2006] provides a thorough overview of the sizeable forecast combination literature.

First, since there is limited support for combining discrete outcomes, this paper considers combining probability forecasts estimated from the probability mass function associated with the one-period-ahead value of random variables. Furthermore, it is shown that probability forecasts from combining discrete choice models instead of combining probability forecasts from discrete choice models, do not yield a discrete choice model with tractable properties. Combination of probability forecasts has already been used in the context of aggregating probability distributions of expert opinions, as discussed in Genest and Zidek [1986] and Clemen and Winkler [1999]. On the methodological side, Kamstra and Kennedy [1998] and Clements and Harvey [2009] among others, investigate different techniques for combining probability forecasts as well as their properties. Second, the paper proposes to use of log and quadratic scoring rules introduced by Brier [1950] and Good [1952], both to evaluate forecasting accuracy of models and to construct adaptive weights for combination. The paper shows that forecast combination achieves greater accuracy in terms of scoring rules. Although probability scoring rules have been applied in the economics and finance literature, for example by Diebold and Rudebusch [1989], Ghysels [1993] and Anderson and Vahid [2001], it has not been employed as a way to combine probability forecasts. Third, the proposed methodology is implemented empirically in forecasting the Fed decisions to change the target rate. The empirical exercise illustrates well that combining probability forecasts improve out-of-sample forecast performance especially when they are combined using weights based on either log or quadratic scoring rules. These results utilise the data and models used by Hu and Phillips [2004a] as a benchmark. The effect of sampling variation around probability forecasts are also assessed through simulations, which follows to some extent the work by McCabe et al. [2011].

The paper is organised as follows. Section 2 sets out the discrete choice model. Forecast combination methodology is discussed in Section 3. Baseline results and robustness checks are provided in Section 4. Section 5 discusses the assessment of the effect of sampling variation associated with probability forecasts. Concluding remarks are presented in Section 6.

Keywords: Forecast combination; Probability forecast; Discrete choice models; Monetary policy decisions.

1 FORECASTING POLICY DECISIONS

1.1 A triple choice model

Following Dueker [1999], the model is

$$r_{t+1}^* = \mathbf{x}'_t \boldsymbol{\beta} - u_{t+1} \tag{1}$$

$$y_{t+1}^* = r_{t+1}^* - r_t \tag{2}$$

where both y_{t+1}^* and r_{t+1}^* are unobservable, \mathbf{x}_t is a $K \times 1$ vector of observable information relevant to the forecast and $\boldsymbol{\beta}$ is a $K \times 1$ vector of parameters. u_{t+1} is an *iid* error with distribution function Φ . If $\Phi(\cdot)$ is the normal (or logistic) distribution with mean zero and variance σ^2 , then the model is the usual ordered probit (or logit). r_{t+1}^* is the optimal policy rate and it is assumed to exist. r_t is the federal funds target rate set by the FOMC in its last meeting. Fed decisions about the target interest rate are classified into three categories: “cut”, “no change” or “hike” the interest rate. Hence,

$$y_{t+1} = \begin{cases} -1 & \text{if } y_{t+1}^* < \mu_1 \\ 0 & \text{if } \mu_1 \leq y_{t+1}^* \leq \mu_2 \\ 1 & \text{if } y_{t+1}^* > \mu_2 \end{cases} \tag{3}$$

is the observed decisions by the Feds. For example, if the difference between the optimal policy rate (r_{t+1}^*) and the actual federal funds target rate (r_t) is greater than the threshold μ_2 then the model would predict a rate hike ($y_{t+1} = 1$). This divergence would need to be substantial to result in a change in the target rate as policy actions are often costly.

In the discrete choice model with error distribution Φ , the probability distribution of y_{t+1} , $\Pr(y_{t+1} = j)$, depends on $(\mathbf{x}_t; \boldsymbol{\theta})$ with $\boldsymbol{\theta} = (\boldsymbol{\beta}', \mu_1, \mu_2, \sigma^2)'$. For simplicity, it is denoted as $P_j(\mathbf{x}_t; \boldsymbol{\theta})$, and is given by

$$\begin{aligned} P_{-1}(\mathbf{x}_t; \boldsymbol{\theta}) &= 1 - \Phi(\mathbf{x}'_t \boldsymbol{\beta} - r_t - \mu_1) \\ P_0(\mathbf{x}_t; \boldsymbol{\theta}) &= \Phi(\mathbf{x}'_t \boldsymbol{\beta} - r_t - \mu_1) - \Phi(\mathbf{x}'_t \boldsymbol{\beta} - r_t - \mu_2) \\ P_1(\mathbf{x}_t; \boldsymbol{\theta}) &= \Phi(\mathbf{x}'_t \boldsymbol{\beta} - r_t - \mu_2) \end{aligned} \tag{4}$$

Hu and Phillips [2004a] defines an indicator function, $\Lambda(t, j)$, where $\Lambda(t, j) = 1$ when $\{y_{t+1} = j\}$. The log-likelihood for the sample $t = 1, \dots, T$ is

$$\log L_T(\boldsymbol{\theta}) = \sum_{t=1}^T \sum_{j=0}^J \Lambda(t, j) \log P_j(\mathbf{x}_t; \boldsymbol{\theta}) \tag{5}$$

Maximising the above log-likelihood for the multiple choice model yields estimates of $\boldsymbol{\theta} = (\boldsymbol{\beta}', \mu_1, \mu_2, \sigma^2)'$. In this paper, the error distribution Φ is specified as a probit following Hu and Phillips [2004a] and Kim et al. [2009]. Furthermore, some of the aggregate macroeconomic time series used as covariates (\mathbf{x}_t) in model (1) by the above mentioned authors, are found to be nonstationary. The same holds for the dataset used in this paper (See section 3 for details). As shown in Hu and Phillips [2004b], standard methods for statistical inference are invalid asymptotically when the estimation process involves integrated processes \mathbf{x}_t .

1.2 Nonstationary data series

Conventional asymptotic theory does not hold for probit or logit regression when variables are nonstationary. Park and Phillips [2000] shows that in the case of binary choice models with nonstationary

explanatory variables, a string of similar decisions (same proportion of unit choices) converges to a random variables that follows an arc sine law with probability density $1/(\pi\sqrt{y(1-y)})$. Park and Phillips also shows that the maximum likelihood estimator is consistent, but it has a dual rate of convergence. Hu and Phillips (2004a, 2004b) and Phillips et al. [2007] extend the results obtained from Park and Phillips [2000] to polychotomous choice models with parametric thresholds determining the choices. These papers propose an expanded range of arc sine laws which allows for diverse distributional shape that permits broken consecutive sequences of similar decisions. The maximum likelihood estimator is also convergent but has either single or dual rate of convergence depending on whether there are only nonstationary variables or a mix of stationary and nonstationary variables in the system.

If the vector \mathbf{x}_t contains integrated processes, then the thresholds are sample size dependent. Hu and Phillips [2004b] scale thresholds by the sample size such that they have the same order of magnitude as the latent variable (y_{t+1}^*) when the information set contains integrated time-series, as such

$$\mu_{iT} = \sqrt{T}\mu_i \quad i = 1, 2$$

Such correction for nonstationarity is not primordial insofar as forecasting accuracy is concerned. It is, however, important when estimating thresholds and conducting statistical inference.

2 FORECAST COMBINATION

In model (1) - (3), the combination problem can be reduced to combining the implied continuous variables, r_{t+1}^* or $P_j(\mathbf{x}_t; \theta)$. As generalised in Timmermann [2006], the aggregator that solves the combination problem maps $N \times 1$ vector of forecast $\hat{\mathbf{z}}_{t+1} \in \mathbb{R}^N$ to a $c \times 1$ summary measure $C(\hat{\mathbf{z}}_{t+1}; \omega_c) \in \mathbb{R}^c \subset \mathbb{R}^N$ with $c < N$. ω_c are the combination parameters of dimension less or equal to N . If that summary measure $C(\hat{\mathbf{z}}_{t+1}; \omega_c) \in \mathbb{R}^1$, then it is a combined point forecast.

The $N \times 1$ vector $\hat{\mathbf{r}}_{t+1}^*$ containing N one-step ahead forecasts of the optimal policy rate can be simply combined as such $C(\hat{\mathbf{r}}_{t+1}^*; \omega_c) = \frac{1}{N} \sum_{i=1}^N \hat{r}_{t+1}^{*(i)}$, $i = 1, \dots, N$ assuming equal weights, $\omega_c = (\frac{1}{N}, \dots, \frac{1}{N})$, for simplification and with $\hat{r}_{t+1}^{*(c)} = C(\hat{\mathbf{r}}_{t+1}^*; \omega_c) \in \mathbb{R}^1$. Hence, the combined point forecast ($\hat{r}_{t+1}^{*(c)}$) implies that equation (2) is now $\hat{y}_{t+1}^{*(c)} = \hat{r}_{t+1}^{*(c)} - r_t$. As discussed in section 1.1, thresholds (μ_1, μ_2) are required to predict the FOMC decisions about the target interest rate. These thresholds cannot be recovered in this combined case, however, as they are nonlinear functions in each model. Even if the thresholds for the combined model were known and fixed, the probabilities of the combination could lie outside the interval given by the individual models.

Proposition 1 *A simple combination of ordered probit models does not yield an ordered probit model with tractable properties.*

Proof: See Appendix A.

Corollary 1 *Proposition 1 also holds for any other monotonic function, such as a Logistic function for example.*

2.1 Combining probability forecasts

Combining probability forecasts provides an intuitive and trivial alternative overcoming the limitations exposed in proposition 1. Although the idea of combining probabilities is not new, the literature has typically focused on specialised issues such as aggregating probability forecasts from expert opinions. Genest and Zidek [1986] addresses the problem of aggregating expert opinions which have been expressed in some numerical form in order to reflect individual uncertainty vis-a-vis a quantity of interest.

Clemen and Winkler [1999] and Genest and Zidek [1986] focus on methods of aggregation with desirable properties rather than accuracy.

Kamstra and Kennedy [1998] concentrates on how to combine individual probability forecasts and/or qualitative forecast, in a computationally attractive manner, through use of logit regression. Kamstra and Kennedy [1998] method encompasses the technique developed by Feather and Kaylen [1989] as a special case. Clements and Harvey [2009] expands the research started by Kamstra and Kennedy [1998] and studies different methods for combining probability forecasts in an optimal manner and the properties of various combination for various data generating processes. The authors argue that the optimal form of combination is unknown as the data generating process of the forecasts and the event being forecast is not known.

In the current setting, at every time period t each model i ($i \in \{1, \dots, N\}$) produces a probability forecast $P_{j,t+1}^{(i)}(\mathbf{x}_t^{(i)}; \boldsymbol{\theta}^{(i)})$ for each state $j = -1, 0, 1$. Hence, the combined one-step ahead probability forecast, $\hat{P}_{t+1}^{(c)}$, simply follows from $\hat{P}_{t+1}^{(c)} = \sum_{i=1}^N \omega_i \hat{P}_{t+1}^{(i)}(\mathbf{x}_t^{(i)}; \hat{\boldsymbol{\theta}}^{(i)})$ where $\hat{P}_{t+1}^{(i)} = (\hat{P}_{-1}^{(i)}, \hat{P}_0^{(i)}, \hat{P}_1^{(i)})'$ is a 3×1 vector with $\hat{\boldsymbol{\theta}}^{(i)}$ as the parameter vector for the i^{th} model. ω_i is a scalar that weights model i . An obvious candidate for ω_i is the simple average, i.e $\omega_i = \frac{1}{N}$. Section 2.3 discusses an optimal way of combining these probability forecasts based on scoring rules as an alternative to the simple average. The next section considers how the forecast accuracy of each model can be assessed.

2.2 Evaluating forecast accuracy with scoring rules

The forecast accuracy can be gauge with scoring rules such as log-score and quadratic score rules. This provides a cleaner way of assessing forecast accuracy rather than matrices counting the amount of correct and incorrect predictions against the actual outcome (for example see table ?? later). Probability scoring rules have been used in the economics and finance literature to evaluate business-cycle turning point and probability forecasts of such turning points (see Diebold and Rudebusch [1989], Ghysels [1993] and Anderson and Vahid [2001] for example). It has also been used to measure statistical accuracy of probability forecasts (see Dawid [1986], Winkler [1993], Winkler [1996], Murphy and Daan [1985]).

Consider the current setting in which a model produces three one-step ahead probability forecasts for each state, namely $(\hat{P}_{-1}, \hat{P}_0, \hat{P}_1)$. If state j happens, then the log-score is given by $S^l = \log(\hat{P}_j)$ similarly to Ng et al. [2010]. Alternatively, forecast accuracy can be assessed with a quadratic-score rule as follows $S^q = 2\hat{P}_j - (\hat{P}_{-1}^2 + \hat{P}_0^2 + \hat{P}_1^2)$

Unlike the log-score rule, the quadratic-score rule proportionately penalises for assigning non-zero probabilities to state(s) that did not occur. Brier [1950] and Good [1952] are the earliest introduction of scoring rules. Lastly, note that when conducting multiple one-step ahead forecasts for each model i over the period (τ_1, τ_2) , the logarithmic or quadratic scores are averaged over the number of forecasted periods

$$\bar{S}_i^q = \frac{1}{\tau_2 - \tau_1} \sum_{t=\tau_1+1}^{\tau_2} S_{it}^q, \quad \bar{S}_i^l = \frac{1}{\tau_2 - \tau_1} \sum_{t=\tau_1+1}^{\tau_2} S_{it}^l$$

where S_{it}^q, S_{it}^l are the quadratic- and log-score obtained for model i at time t .

2.3 Score based weights

Instead of combining forecasts with a simple average, combination can be derived from each model's forecasting performance. This paper proposes two intuitive symmetric weighting methods constructed from scoring rules, both log-score and quadratic-score rules.

Definition 1 *The score based weights are defined as*

$$\omega_i^q = \frac{\bar{S}_i^q}{\sum_{i=1}^N \bar{S}_i^q}, \quad \omega_i^l = \frac{1/|\bar{S}_i^l|}{\sum_{i=1}^N 1/|\bar{S}_i^l|} \quad i = 1, \dots, N$$

where ω_i^q, ω_i^l are the weight for forecast i based on the average quadratic- (\bar{S}_i^q) and log-score (\bar{S}_i^l) respectively.

The $1/|\cdot|$ in log-score weights (ω_i^l) is constructed as such because the log-scoring rule returns negative values. From the definition, it is clear that the highest the scoring rule for a forecasting model, the higher the weight given to its one-step ahead forecast. Furthermore, the composition of the weights changes over time as the scores are averaged.

3 RESULTS

3.1 Out-of-sample forecast

The out-of-sample forecast is conducted using a recursive forecasting scenario. The sample is cut in half such that the estimation sample spans over $\{1, \dots, \lfloor T/2 \rfloor\}$ and the one-step-ahead forecast runs in the sample spanning over $\{\lfloor T/2 \rfloor + 1, \dots, T\}$.¹ Once a month is forecasted, the estimation window is increased by adding an observation to the sample, and so on until the end of the sample. Out-of-sample forecasting ability is evaluated for two multivariate models (with 4 and 10 variables) and six combination models: equal weights combination, log weights combination and quadratic weight combination (with 4 and 10 variables).

The out-of-sample forecast results are presented in Table 1. Both scoring rules indicate that the quadratic score weighted combination outperforms all other models for most time periods, whether combined or composed of multiple variables. For the period from 1994 to 2001, however, the log score and quadratic score are in disaccord. The former favours the quadratic score weighted combination whereas the latter picks the 4 variables model. Two predictors, annual growth for manufacturer's new orders and initial claims for unemployment insurance, appear to have higher scores than the rest across the different time sample. These two predictors are among the 4 variables selected by Hu and Phillips [2004a]. Note that the negative quadratic scores are set to zero when computing the combination weights based on quadratic scoring rules.

4 CONCLUDING COMMENTS

This paper provides a methodology to combine forecasts based on several discrete choice models. It shows that a simple combination of ordered probit models does not yield an ordered probit model with tractable properties. Instead, it is easier and more intuitive to combine probability forecasts. Log-scoring rules and quadratic-scoring rules are used both to evaluate the forecasting accuracy of each model and to combine the probability forecasts. In addition to producing point forecasts, the effect of sampling variation is also assessed through simulations. Forecast combination is applied to forecast the Federal Open Market Committee (FOMC) decisions in changing the federal funds target rate. The empirical results show that combining forecasted probabilities using quadratic and log scoring rules mostly outperforms both equal weight combination and forecasts based on multivariate models.

A APPENDIX: PROOF

Available upon request.

¹ $\lfloor \cdot \rfloor$ is a floor function for integer round up.

Table 1: Out-of-Sample Forecasts (Recursive Forecasts)

Models\Scores	1994-2001		1994-2008		1994-2010	
	Log	Quadratic	Log	Quadratic	Log	Quadratic
Baseline of no change	-	0	-	0	-	0.19
Multivariate Models & Forecast Combinations						
Multivariate (10 var)	-1.43	0.46	-1.36	0.34	-1.79	0.22
Multivariate (4 var)	-1.00	0.57	-0.95	0.47	-1.40	0.33
Equal weights (10 var)	-1.23	0.29	-1.09	0.32	-1.02	0.36
Equal weights (4 var)	-1.00	0.40	-0.88	0.45	-0.86	0.46
Log weights (10 var)	-1.19	0.31	-0.88	0.46	-0.93	0.43
Log weights (4 var)	-1.00	0.39	-0.79	0.51	-0.83	0.48
Quadratic weights (10 var)	-1.18	0.31	-0.81	0.50	-0.87	0.46
Quadratic weights (4 var)	-0.99	0.40	-0.78	0.52	-0.83	0.48
Univariate Models						
M2	-1.31	0.27	-1.04	0.37	-1.04	0.36
Unemployment claim	-1.09	0.40	-0.90	0.50	-1.10	0.44
Consumer Confidence	-2.22	0.18	-1.79	0.13	-1.12	0.33
New orders	-0.83	0.51	-0.84	0.48	-0.93	0.40
Total Capacity Util.	-1.82	0.09	-1.05	0.36	-0.93	0.43
Unemployment rate	-2.40	0.12	-1.71	0.13	-1.26	0.25
NAPM	-1.34	0.23	-1.73	0.17	-1.68	0.15
Average Working Hrs	-2.86	0.06	-2.51	-0.07	-1.73	0.07
Industrial Prod.	-2.83	0.12	-2.20	-0.02	-1.62	-0.01
Core CPI	-2.77	0.08	-2.02	0.05	-1.50	0.10

Notes: The numbers in the table are the log and quadratic scoring rules as introduced in section 2.2. *Baseline of no change* model corresponds to predicting no changes in the interest rate throughout the forecasting sample. *Multivariate (10 var)* is a model with all 10 regressors. *Multivariate (4 var)* is the model selected by Hu and Phillips [2004a] containing initial claim for unemployment, annual growth of M2, Consumer Confidence, annual growth of manufacturers' new orders (Appendix B describes the data). *Equal weights (10 var)* combines the probability forecasts of 10 univariate models featuring all 10 variables. *Equal weights (4 var)* combines 4 univariate models for the same 4 variables as in *Multivariate (4 var)*. *Log weights* and *Quadratic weights* with 10 and 4 variables refer to the models combining probability forecasts following definition 1 in section 2.2. The best performing models in each subcategory are shaded.

B APPENDIX: DATA

Available upon request.

REFERENCES

- Anderson, H. and F. Vahid (2001). Predicting the probability of a recession with nonlinear autoregressive leading indicator models. *Macroeconomic Dynamics* 5, 482–505.
- Bates, J. M. and C. W. J. Granger (1969). The combination of forecasts. *Operations Research Quarterly* 20, 451–468.
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review* 75, 1–3.
- Clemen, R. T. and R. L. Winkler (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis* 19, 187–203.
- Clements, M. P. and D. I. Harvey (2009, August). Combining probability forecasts. Mimeo.

- Dawid, A. P. (1986). Probability forecasting. In S. Kotz, N. L. Johnson, and C. B. Read (Eds.), *Encyclopedia of Statistical Sciences*, Volume 7, pp. 210–218. John Wiley and Sons.
- Diebold, F. X. and G. D. Rudebusch (1989). Scoring the leading indicators. *Journal of Business* 62, 369–391.
- Dueker, M. (1999). Measuring monetary policy inertia in target fed funds rate changes. *Federal Reserve Bank of St. Louis Review* 81(5), 3–9.
- Feather, P. M. and M. S. Kaylen (1989). Conditional qualitative forecasting. *American Journal of Agricultural Economics* 71(1), 195–201.
- Genest, C. and J. V. Zidek (1986). Combining probability distributions: A critique and an annotated bibliography. *Statistical Science* 1, 114–148.
- Ghysels (1993). On scoring asymmetric periodic probability models of turning-point forecasts. *Journal of Forecasting* 12(227-238).
- Good, I. (1952). Rational decisions. *Journal of the Royal Statistical Society. Series B* 14(1), 107–114.
- Hu, L. and P. C. B. Phillips (2004a). Dynamics of the federal funds target rate: A nonstationary discrete choice approach. *Journal of Applied Econometrics* 19(7), 851–867.
- Hu, L. and P. C. B. Phillips (2004b). Nonstationary discrete choice. *Journal of Econometrics* 120(1), 103–138.
- Kamstra, M. and P. Kennedy (1998). Combining qualitative forecasts using logit. *International Journal of Forecasting* 14, 83–93.
- Kim, H., J. Jackson, and R. Saba (2009). Forecasting the fomc’s interest rate setting behavior: A further analysis. *Journal of Forecasting* 28, 145–165.
- McCabe, B. P., G. M. Martin, and D. Harris (2011). Efficient probabilistic forecasts for counts. *Journal of the Royal Statistical Society. Series B* 73(3), 1–20.
- Murphy, A. H. and H. Daan (1985). Forecast evaluation. In A. H. Murphy and R. W. Katz (Eds.), *Probability, Statistics and Decision Making in the Atmospheric Sciences*, pp. 379–437. Westview Press.
- Ng, J., C. S. Forbes, G. M. Martin, and B. P. McCabe (2010, October). Non-parametric estimation of forecast distributions in non-gaussian state space models. Monash University, Mimeo.
- Park, J. Y. and P. C. B. Phillips (2000). Nonstationary binary choice. *Econometrica* 68(3), 1249–1280.
- Phillips, P. C., S. Jin, and L. Hu (2007, December). Nonstationary discrete choice: A corrigendum and addendum. *Journal of Econometrics* 141(2), 1115–1130.
- Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics* 24, pp. 135–196. Elsevier, North-Holland.
- Winkler, R. L. (1993). Evaluating probabilities: asymmetric scoring. *Management Science* 4, 1395–1405.
- Winkler, R. L. (1996). Scoring rules and the evaluation of probabilities (with discussion). *Test* 5(1), 1–60.