Evaluating point and density forecasts of DSGE models: Appendix

Maik H. Wolters *
University of Kiel and Kiel Institute for the World Economy

September 5, 2013

Appendix A: Greenbook forecasts from different data sources

I use the dataset by Faust and Wright (2009) from Jon Faust's website to compare the Greenbook forecasts reported in this dataset to those reported in a dataset that is available from the Federal Reserve Bank of Philadelphia. The latter contains the scanned original Greenbook documents.[1] Using the same definition of revised data as Faust and Wright (2009) I can replicate the Greenbook RMSEs in the paper exactly.[2] However, when plotting forecast errors, I found some surprisingly large values, which do not occur when basing the analysis on the Philadelphia Fed Greenbook data.

Output growth forecasts

Table 1 lists all discrepancies between the Greenbook output growth forecasts from the Faust and Wright (2009) dataset and the one available from the Federal Reserve Bank of Philadelphia that exceed 0.2%. The table is based on data from 1984-2000, i.e. the sample that is used for the main results of the paper. Some of the discrepancies might be attributed to the fact that Faust and Wright (2009) compute approximate growth rates as $400\log(x_t/x_{t-1})$, rather than computing accurate growth rates as $100[(x_t/x_{t-1})^4-1]$. Based on these computations Faust and Wright (2009) report series with many decimal places, while the Philadelphia Fed figures are reported up to the first decimal place.

*Mailing address: Kiel Institute for the World Economy, Hindenburgufer 66, 24105 Kiel, Germany; maik.wolters@ifw-kiel.de.

[1]The dataset used by Faust and Wright (2009) is available on http://e105.org/faustj/download/faustWrightGBTSdata.zip?d=n. Greenbook forecasts are also available from the Federal Reserve Bank of Philadelphia in two datasets: forecasts that are closest to the middle of a quarter are available in an Excel sheet and all other forecasts are available as .pdf documents that contain scanned original Greenbook documents. The URLs are: http://www.philadelphiafed.org/research-and-data/real-time-center/greenbook-data/philadelphia-data-set.cfm and http://www.philadelphiafed.org/research-and-data/real-time-center/greenbook-data/pdf-data-set.cfm.

[2]They use the data vintage that was released two quarters after the quarter to which the data refer to evaluate the forecasting accuracy. For example, revised data for 1999Q1 is obtained by selecting the entry for 1999Q1 from the data vintage released in 1999Q3. They use the Federal Reserve Bank of Philadelphia's real-time dataset. It is available on: http://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data.

However, some of the discrepancies cannot be explained by mere rounding errors and it remains unclear from where the differences arise. For the Greenbook from May 15, 1997, GDP growth has based on the Faust & Wright data a forecast error (actual minus predicted) of -9.58% for the nowcast and of 11.59% for one quarter ahead. The corresponding annualized GDP growth forecasts are 12.84% for the nowcast and -8.56% for one quarter ahead. Revised data shows entries of 3.25% and 3.03%, respectively.

Table 1: Greenbook output growth forecasts from two data sources

date	Faust & Wright	Philadelphia Fed	difference
	Horizon: 0		
19840321	7.6591	8.0	-0.3409
19970515	12.8371	1.8	11.0371
19990128	3.6747	2.7	0.9747
	Horizon: 1		
19840321	5.7972	6.0	-0.2028
19881207	5.5942	5.8	-0.2058
19970515	-8.5586	2.5	-11.0586
	Horizon: 5		
19910925	7.1815	3.1	4.0815

Notes: the table shows annualized output growth forecasts from the Greenbook as reported in the dataset used by Faust and Wright (2009) and the Greenbook forecasts available from the Federal Reserve Bank of Philadelphia. The table only reports discrepancies of 0.2% or higher.

To check how these differences affect the assessment of the forecasting accuracy, I compute RMSEs for the Philadelphia Fed dataset and compare them to the results by Faust and Wright (2009). Table 2 shows that the dataset by Faust and Wright (2009) yields a lower forecasting accuracy for horizons 0 and 1 than the dataset by the Philadelphia Fed.

Table 2: Greenbook output growth RMSEs based on two different data sources

horizon	Faust & Wright	Philadelphia Fed
0	1.75	1.52
1	2.12	1.85
2	2.01	2.01
3	2.15	2.15
4	2.08	2.08
5	2.08	2.08

Notes: The RMSEs are computed based on all available forecasts in the dataset from Faust and Wright (2009) from 1984-2000 and the corresponding data from the Federal Reserve Bank of Philadelphia.

To assess whether these differences might change the interpretation of the main results in Faust and Wright (2009), table 3 replicates the results for horizons 0 and 1 from table 2 in Faust and Wright (2009) when using the Philadelphia Fed's Greenbook forecasts (Phil Fed) compared to using the dataset from Faust and Wright (F&W). To compute the values for the Philadelphia Fed dataset one needs to multiply for each forecasting model the relative RMSE from Faust and Wright (2009) with the Greenbook RMSE to get the absolute RMSE. Afterwards I divide the absolute RMSE with the Greenbook RMSE from the Philadelphia Fed dataset to get the new relative RMSE. Of course, I cannot say anything about the statistical significance in the difference of the forecasting accuracy of models relative to the Greenbook.

Table 3: Greenbook RMSE and relative RMSE of time series models: 1984-2000

Output growth

horizon	GB	RAR	DAR	EWA	BMA	FAA	FAV	DF	FVS
				jump off -1					
0 (F&W)	1.75	1.09	1.09	1.09	1.10	1.30	1.39	1.32	1.24
0 (Phil Fed)	1.52	1.25	1.25	1.25	1.27	1.50	1.60	1.52	1.43
1 (F&W)	2.12	**0.87**	**0.86**	**0.87**	**0.93**	1.13	1.20	1.15	0.93
1 (Phil Fed)	1.85	1.00	**0.99**	1.00	1.07	1.29	1.38	1.32	1.07
				jump off 0					
1 (F&W)	2.12	**0.84**	**0.84**	**0.84**	**0.85**	**0.96**	1.07	**0.99**	**0.85**
1 (Phil Fed)	1.85	**0.96**	**0.96**	**0.96**	**0.97**	1.10	1.23	1.13	**0.97**

Notes: GB: Greenbook; RAR: recursive autoregression; DAR: direct forecast from autoregression; EWA: equal-weighted averagin; BMA: Bayesian model averaging; FAA: factor augmented autoregression; FAV: factor augmented vector autoregression; DF: dynamic factor model; FVS: factor-spanned variable selection. A detailed description of the different models is contained in Faust and Wright (2009).

Faust and Wright (2009) stress that the Greenbook nowcast is extremely accurate compared to other methods due to the fact that the Fed makes great efforts to mirrorring key elements of the data construction process of the BEA. Using the Philadelphia Fed's dataset even increases the accuracy of the Greenbook nowcast to a RMSE of 1.52 compared to 1.75 in Faust and Wright (2009). They use a forecast from a simple autoregressive process (RAR and DAR) as a benchmark and find that this is at least as accurate as the other considered time series models. The results in table 3 show that the relative RMSE of the autoregressive process is only 1.25 when using the Philadelphia Fed dataset compared to 1.09 when using the Faust & Wright dataset. This stresses once more the extremely high accuracy of the Greenbook nowcasts.

For the one quarter ahead output growth forecasts the results by Faust and Wright (2009) give the impression that many simple time series methods yield a better forecast than the Greenbook even though the difference is statistically not significant. Using the Philadelphia Fed dataset makes clear that the best forecasts from time series models are as good as the Greenbook forecasts, but not better. The best relative RMSE is 0.99 for the direct autoregressive forecast. In contrast, using the Faust and Wright (2009) dataset one gets a relative RMSE of 0.86. To sum up, the main findings of the paper regarding the output growth forecasts are not affected. However, the dataset by Faust and Wright (2009) understates the Greenbook's forecasting accuracy compared to the autoregressive forecast to some extent.

Inflation forecasts

While I did not find any extremely large forecasting errors for inflation, the inflation forecasts from the two dataset differ from each other. Table 4 shows discrepancies that exceed 0.2% for the sample 1984-2000.[3]

Table 4: Greenbook inflation forecasts from two data sources

date	Faust & Wright	Philadelphia Fed	difference
	Horizon: 0		
19960626	2.0783	1.6	0.4783
19960918	2.0783	1.7	0.3783
19961212	2.4693	2.2	0.2693
	Horizon: 1		
19911030	3.0529	3.7	-0.6471
19960626	2.5668	2.2	0.3668
19960918	2.5668	2.3	0.2668
	Horizon: 2		
19960626	2.6642	2.4	0.2642
	Horizon: 3		
19860813	2.5668	2.8	-0.2332
19960626	2.7615	2.3	0.4615
19961212	2.2739	2.0	0.2739
	Horizon: 4		
19960626	2.7615	2.3	0.4615
19960918	2.5668	2.3	0.2668
19961212	2.2739	2.0	0.2739
	Horizon: 5		
19960626	2.7615	2.3	0.4615
19960918	2.6642	2.4	0.2642
19961212	2.5668	2.3	0.2668

Notes: the table shows annualized inflation forecasts from the Greenbook as reported in the dataset used by Faust and Wright (2009) and the Greenbook forecasts available from the Federal Reserve Bank of Philadelphia. The table only reports discrepancies of 0.2% or higher.

Most of the larger discrepancies originate from Greenbook forecasts from 1996. In 1996 the Fed changed from using the MPS model to the FRB/US model to compute forecasts. However,

[3]The difference of the inflation forecasts between 1980 and 1984 exceeds 0.2% in most cases. The inflation forecasts from the Faust& Wright dataset are for these observations with only few exceptions lower than the forecasts from the Philadelphia Fed dataset. However, these observations are not used for the main results of Faust and Wright (2009).

it is unclear whether this change is a reason for some of the larger discrepancies between the two datasources. As there are no extreme discrepancies as in the case of output growth, all the inflation forecast results in Faust and Wright (2009) are robust to using the dataset from the Federal Reserve Bank of Phildelphia. Table 5 shows RMSE from the two datasets. The differences are small.

Table 5: Greenbook inflation RMSEs based on two different data sources

horizon	Faust & Wright	Philadelphia Fed
0	0.69	0.70
1	0.79	0.80
2	0.81	0.82
3	0.93	0.94
4	0.89	0.91
5	1.14	1.15

Notes: The RMSEs are computed based on all available forecasts in the dataset from Faust and Wright (2009) from 1984-2000 and the corresponding data from the Federal Reserve Bank of Philadelphia.

Appendix B: Additional Results

An additional forecasting accuracy measure

Faust and Wright (2009) present a table showing the percentage of forecast periods in which the non-structural forecasts are more accurate than the Greenbook projections. This metric is not as sensitive to outliers as the RMSEs. I compute accordant numbers for the DSGE models which are shown in the table 6. This metric leads to the same relative rankings of the forecasting models as in table 2 in the paper.

Table 6: Percentage of periods alternative forecast better than Greenbook: 1984-2000

(a) Output growth

horizon	DS	FM	SW	EDO	best FW	worst FW
			jump off -1			
0	30	36	36	34	43	30
1	50	47	47	48	**60**	39
2	47	46	**53**	50	**58**	37
3	46	43	**59**	**51**	**57**	42
4	43	45	**52**	46	**54**	36
5	42	43	**59**	48	**52**	43
			jump off 0			
1	43	48	48	48	**59**	40
2	50	50	**58**	42	**55**	41
3	47	48	**57**	49	**57**	38
4	43	46	**54**	42	**57**	39
5	44	43	**54**	43	49	43

(b) Inflation

horizon	DS	FM	SW	EDO	best FW	worst FW
			jump off -1			
0	43	31	43	30	37	25
1	30	30	42	38	40	21
2	41	35	37	35	38	25
3	46	38	37	30	39	17
4	43	30	36	31	43	11
5	34	30	38	34	46	16
			jump off 0			
1	37	31	36	44	41	30
2	37	33	39	46	40	21
3	41	42	39	39	43	20
4	39	25	32	39	43	18
5	39	30	33	**53**	48	15

(c) Federal Funds Rate

horizon	DS	FM	SW	EDO	best FW	worst FW
			jump off -1			
0	8	12	7	5	-	-
1	29	27	21	11	-	-
2	44	32	32	18	-	-
3	49	34	40	24	-	-
4	**56**	31	46	30	-	-
5	**59**	34	50	29	-	-
			jump off 0			
1	33	30	30	24	-	-
2	42	36	40	27	-	-
3	48	41	48	26	-	-
4	48	39	**53**	28	-	-
5	**52**	42	**54**	25	-	-

Notes: The first column shows the forecast horizon. The other columns show the percentage of forecast periods in which forecast errors of specific models are smaller in absolute value than the Greenbook forecast error. Entries greater than 50 percent indicate that the alternative forecast is better more than half the time and are in bold. The results for bestFW and worstFW are only indicative (and probably too high for output growth horizons 0 and 1) as they are based on another Greenbook dataset (see Appendix A).

Details on the computation of combination weights

Let I_t^m be the information set of model m at time t including the model equations, parameter estimates and the observable time series of the accordant data vintage. A combined point forecast of models $m = 1,...,M$ for horizon h denoted as $E[y_{t+h}^{obs}|I_t^1,...,I_t^M,\omega_{1,h},...,\omega_{M,h}]$ can be written as the weighted sum of individual density forecasts $p[y_{t+h}^{obs}|I_t^m]$ with assigned weights $\omega_{m,h}$ divided by the number of draws S:

$$E[y_{t+h}^{obs}|I_t^1,...,I_t^M,\omega_{1,h},...,\omega_{M,h}] = \frac{1}{S}\sum_{m=1}^{M}\omega_{m,h}p[y_{t+h}^{obs}|I_t^m]. \tag{1}$$

In the following, I discuss various methods how to choose the weights $\omega_{m,h}$.

A natural way to weight different models in a Bayesian context is to use Bayesian Model Averaging. The marginal likelihood $ML(y_T^{obs}|m)$—with T denoting all observations of a specific historical data sample observed in period t—is computed for each model $m = 1,...,M$ and posterior probability weights are given by:

$$\omega_m = ML(m|y_T^{obs}) = \frac{ML(y_T^{obs}|m)}{\sum_{m=1}^{M} ML(y_T^{obs}|m)}, \tag{2}$$

where a flat prior belief about model m being the true model is used so that no prior beliefs show up in the formula. This weighting scheme is based on the fit of a model to the observed time series. Unfortunately, posterior probability weights are not comparable for models that are estimated on a different number of time series. A second problem of the posterior probability weights is that over-parameterized models that have an extreme good in-sample fit, but a bad out-of-sample forecasting accuracy, are assigned high weights. To circumvent these problems I use an out-of-sample weighting scheme based on predictive likelihoods as proposed by Eklund and Karlsson (2007) and Andersson and Karlsson (2007) as described in the following paragraph.

Predictive Likelihood (PL) The available data is split into a training sample used to estimate the models and a hold-out sample used to evaluate each model's forecasting performance. The forecasting performance is measured by the predictive likelihood, i.e. the marginal likelihood of the hold-out sample conditional on a specific model. I follow the approach suggested by Andersson and Karlsson (2007) to compute a series of small hold-out sample predictive likelihoods for each horizon. Equation (3) shows how to compute the predictive likelihood PL of model m for horizon h:

$$PL_h^m = ML(y_{holdout}^{obs}|y_{training}^{obs}) = \prod_{t=l}^{T-h} ML(y_{t+h}^{obs}|y_t^{obs}). \tag{3}$$

Starting with an initial trainings sample of length l, one computes the marginal likelihood for horizon h using the hold-out sample. The training sample is expanded by one observation to $l+1$ and a second maginal likelihood is computed for the hold-out sample that is one observation shorter than the previous one. This continues until the trainings sample has increased to lenght $T-h$ and the hold-out sample has shrinked to length h. To make the results comparable among models, only the three common variables output growth, inflation and the interest rate are considered for the computation of the predictive likelihood. Finally, the predictive likelihood weights are computed by replacing the marginal likelihood in equation (2) with the predictive likelihood:

$$\omega_{m,h} = \frac{PL_h^m}{\sum_{m=1}^{M} PL_h^m}. \tag{4}$$

The predictive likelihood weighting scheme allows for different weights to be assigned to a given model at different forecast horizons.

Ordinary Least Squares Weights (OLS) In model averaging applications of time series models it is common to assume a linear-in-weights model and estimate combination weights by ordinary least squares (see Timmermann, 2006). I use the forecasts from previous vintages for each model and the accordant data realizations to regress the realizations y^{obs}_{t+h} on the forecasts $E[y^{obs}_{t+h}|I^m_t]$ from the different models via constrained OLS separately for each variable:

$$y^{obs}_{t+h} = \omega_{1,h}E[y^{obs}_{t+h}|I^1_t] + ... + \omega_{M,h}E[y^{obs}_{t+h}|I^M_t] + \varepsilon_{t+h}, \qquad s.t. \sum_{m=1}^{M} \omega_{m,h} = 1. \tag{5}$$

The resulting parameter estimates $\omega_{1,h},...,\omega_{M,h}$ are the combination weights. Therefore, the combination weights differ for different horizons and also for the three different variables. I omit an intercept term and restrict the weights to sum to one so that the weights can be interpreted as the fractions the specific models contribute to the weighted forecast. It also ensures that the combined forecast lies inside the range of the individual forecasts.

RMSE based weights (RMSE) There are several ways to compute simple relative performance weights. I consider here weightings based on RMSEs of past forecasts and weights based on the relative past forecasting accuracy by ranking the accuracy of the different models. For the prior case RMSE based weights can be computed by taking forecasts from previous vintages and compute the RMSE for each model. The weights are then calculated by taking the inverse relative RMSE performance:

$$\omega_{m,h} = \frac{(1/RMSE^m_h)}{\sum_{m=1}^{M}(1/RMSE^m_h)}. \tag{6}$$

Rank based weights (Rank) A second possibility to compute relative performance weights is to assign ranks R from 1 to M according to the past forecasting accuracy measured by the RMSEs. This method is similar to the RMSE based weights while being more robust to outliers. The performance rank based weights are computed as follows:

$$\omega_{m,h} = \frac{(1/R^m_h)}{\sum_{m=1}^{M}(1/R^m_h)}. \tag{7}$$

Both methods can assign different weights to forecasts of different variables and different forecasting horizons.

Mean Forecast (Mean) The simplest method to compute a weighted forecast is to give equal weight to each model and simply compute the mean forecast of all models. From a theoretical point of view this approach is not preferable as the weights are purely subjective prior weights implicitly given by the choice of models. However, it has often been found that simple weighting schemes perform well (see e.g. Hsiao et al., 2010). A reason is that they give weight to several models instead of choosing one optimal model and are thus robust.

Median Forecast (Median) Another possibility is to choose the median of different model forecasts. I compute the median forecast for each of the ordered draws of all models. This gives the density of the median forecast which is used to compute the mean of all these draws as a point forecast. The approach is similar to taking the mean forecast, but is more robust to outliers. The medians from the ordered forecast draws need not to come from the same model for different slices of the ordered forecast draws. By counting the fraction that the median forecast is generated by a specific model one can compute pseudo weights of the different model forecasts that show the contribution of each model to the final point forecast.

Example of weights given to different forecasting models

The results in the paper show that model combination methods that give weight to several models perform well. Likelihood based weighting methods are preferable in theory, but do not work as well in practice. Differences in predictive likelihoods of different models are so high that at most times all weight is given to a single model. Tables 7 to 9 report as an example model weights for forecasts derived from data vintage May 12, 2000. These show that combined forecasts based on the predictive likelihood and OLS-weights tend to give all weight to one model rather than averaging over the forecasts from several models.

Table 7: Combination weights for data vintage May 12, 2000: output growth

model	PL	OLS	Median	Mean	RMSE	Rank
			horizon 0			
DS	1.00	0.00	0.50	0.25	0.24	0.12
FM	0.00	1.00	0.21	0.25	0.26	0.48
SW	0.00	0.00	0.29	0.25	0.24	0.16
EDO	0.00	0.00	0.00	0.25	0.26	0.24
			horizon 1			
DS	1.00	0.00	0.48	0.25	0.24	0.16
FM	0.00	0.00	0.02	0.25	0.23	0.12
SW	0.00	0.50	0.50	0.25	0.27	0.24
EDO	0.00	0.50	0.00	0.25	0.26	0.48
			horizon 2			
DS	1.00	0.00	0.00	0.25	0.24	0.16
FM	0.00	0.06	0.50	0.25	0.23	0.12
SW	0.00	0.44	0.49	0.25	0.27	0.24
EDO	0.00	0.50	0.01	0.25	0.26	0.48
			horizon 3			
DS	1.00	0.00	0.00	0.25	0.24	0.16
FM	0.00	0.12	0.50	0.25	0.23	0.12
SW	0.00	0.36	0.47	0.25	0.26	0.48
EDO	0.00	0.52	0.03	0.25	0.27	0.24
			horizon 4			
DS	1.00	0.00	0.00	0.25	0.24	0.12
FM	0.00	0.29	0.50	0.25	0.24	0.16
SW	0.00	0.22	0.45	0.25	0.26	0.48
EDO	0.00	0.49	0.05	0.25	0.26	0.24
			horizon 5			
DS	1.00	0.00	0.00	0.25	0.23	0.12
FM	0.00	0.41	0.50	0.25	0.26	0.24
SW	0.00	0.34	0.38	0.25	0.27	0.48
EDO	0.00	0.25	0.12	0.25	0.24	0.16

Notes: PL: Predictive Likelihood; OLS: Ordinay Least Squares; Median: Median forecast; Mean: Mean forecast; RMSE: weighted by inverse RMSE; Rank: weighted by inverse ranks; DS: Del Negro & Schorfheide; FM: Fuhrer & Moore; SW: Smets & Wouters; EDO: FRB/EDO Model by Edge, Kiley & Laforte; The first column shows the model name and the rows show the weight of each model for the different combination schemes. For each horizon, the four model weights sum up to 1.

Table 8: Combination weights for data vintage May 12, 2000: inflation

model	PL	OLS	Median	Mean	RMSE	Rank
			horizon 0			
DS	1.00	0.21	0.01	0.25	0.28	0.24
FM	0.00	0.05	0.50	0.25	0.20	0.12
SW	0.00	0.74	0.49	0.25	0.29	0.48
EDO	0.00	0.00	0.00	0.25	0.23	0.16
			horizon 1			
DS	1.00	0.00	0.29	0.25	0.26	0.24
FM	0.00	0.00	0.21	0.25	0.22	0.12
SW	0.00	0.88	0.50	0.25	0.28	0.48
EDO	0.00	0.12	0.00	0.25	0.24	0.16
			horizon 2			
DS	1.00	0.00	0.26	0.25	0.26	0.24
FM	0.00	0.35	0.24	0.25	0.24	0.16
SW	0.00	0.40	0.50	0.25	0.28	0.48
EDO	0.00	0.25	0.00	0.25	0.22	0.12
			horizon 3			
DS	1.00	0.16	0.29	0.25	0.29	0.48
FM	0.00	0.34	0.21	0.25	0.21	0.12
SW	0.00	0.09	0.50	0.25	0.28	0.24
EDO	0.00	0.41	0.00	0.25	0.22	0.16
			horizon 4			
DS	1.00	0.00	0.30	0.25	0.28	0.24
FM	0.00	0.30	0.20	0.25	0.19	0.12
SW	0.00	0.16	0.50	0.25	0.28	0.48
EDO	0.00	0.54	0.00	0.25	0.25	0.16
			horizon 5			
DS	1.00	0.00	0.33	0.25	0.27	0.24
FM	0.00	0.32	0.17	0.25	0.20	0.12
SW	0.00	0.16	0.50	0.25	0.28	0.48
EDO	0.00	0.52	0.00	0.25	0.25	0.16

Notes: PL: Predictive Likelihood; OLS: Ordinay Least Squares; Median: Median forecast; Mean: Mean forecast; RMSE: weighted by inverse RMSE; Rank: weighted by inverse ranks; DS: Del Negro & Schorfheide; FM: Fuhrer & Moore; SW: Smets & Wouters; EDO: FRB/EDO Model by Edge, Kiley & Laforte; The first column shows the model name and the rows show the weight of each model for the different combination schemes. For each horizon, the four model weights sum up to 1.

Table 9: Combination weights for data vintage May 12, 2000: Federal Funds Rate

model	PL	OLS	Median	Mean	RMSE	Rank
			horizon 0			
DS	1.00	0.00	0.44	0.25	0.24	0.16
FM	0.00	0.00	0.50	0.25	0.28	0.24
SW	0.00	0.00	0.06	0.25	0.30	0.48
EDO	0.00	1.00	0.00	0.25	0.18	0.12
			horizon 1			
DS	1.00	0.00	0.02	0.25	0.24	0.16
FM	0.00	0.00	0.50	0.25	0.31	0.48
SW	0.00	0.00	0.48	0.25	0.27	0.24
EDO	0.00	1.00	0.00	0.25	0.18	0.12
			horizon 2			
DS	1.00	0.00	0.03	0.25	0.25	0.16
FM	0.00	0.00	0.50	0.25	0.29	0.48
SW	0.00	0.00	0.47	0.25	0.26	0.24
EDO	0.00	1.00	0.00	0.25	0.20	0.12
			horizon 3			
DS	1.00	0.00	0.03	0.25	0.26	0.16
FM	0.00	0.00	0.50	0.25	0.27	0.48
SW	0.00	0.00	0.47	0.25	0.27	0.24
EDO	0.00	1.00	0.00	0.25	0.20	0.12
			horizon 4			
DS	1.00	0.00	0.04	0.25	0.27	0.24
FM	0.00	0.00	0.50	0.25	0.25	0.16
SW	0.00	0.00	0.46	0.25	0.27	0.48
EDO	0.00	1.00	0.00	0.25	0.21	0.12
			horizon 5			
DS	1.00	0.00	0.04	0.25	0.27	0.24
FM	0.00	0.00	0.50	0.25	0.22	0.12
SW	0.00	0.00	0.46	0.25	0.29	0.48
EDO	0.00	1.00	0.00	0.25	0.22	0.16

Notes: PL: Predictive Likelihood; OLS: Ordinay Least Squares; Median: Median forecast; Mean: Mean forecast; RMSE: weighted by inverse RMSE; Rank: weighted by inverse ranks; DS: Del Negro & Schorfheide; FM: Fuhrer & Moore; SW: Smets & Wouters; EDO: FRB/EDO Model by Edge, Kiley & Laforte; The first column shows the model name and the rows show the weight of each model for the different combination schemes. For each horizon, the four model weights sum up to 1.