Iterated Multi-Step Forecasting with Model Coefficients Changing Across Iterations

Michal Franta[1]

Abstract:

Iterated multi-step forecasts are usually constructed assuming the same model in each forecasting iteration. In this paper, the model coefficients are allowed to change across forecasting iterations according to the in-sample prediction performance at a particular forecasting horizon. The technique can thus be viewed as a combination of iterated and direct forecasting. The superior point and density forecasting performance of this approach is demonstrated on a standard medium-scale vector autoregression employing variables used in the Smets and Wouters (2007) model of the US economy. The estimation of the model and forecasting are carried out in a Bayesian way on data covering the period 1959Q1–2016Q1.

JEL Codes: C11; C32; C53;

Keywords: Multi-step forecasts, VAR, Bayesian estimation, iterated forecasting, direct forecasting, density forecasting

[1] Michal Franta, Czech National Bank, e-mail: michal.franta@cnb.cz
I would like to thank Jan Bruha, Andrea Carriero, Simona Malovana, Howell Tong, and seminar participants at the Czech National Bank for useful comments. The views expressed here are those of the author and not necessarily those of the Czech National Bank.

1. Introduction

Multi-step forecasting is one of the most important tasks in applied macroeconomics. Several approaches have emerged, and their relative accuracy often depends on the forecasting technique and the circumstances of the forecasting exercise. In this paper the focus is primarily on iterated multi-step forecasting, i.e., on a forecasting technique that combines one-step-ahead forecasts into forecasts for several periods ahead. Iterated forecasts are sometimes referred to as plug-in forecasts, reflecting the sequential nature of the forecasting process.

Models used for forecasting are often estimated by means of the likelihood function. Maximum likelihood estimation effectively means that the parameter estimates minimize the one-step-ahead prediction errors within the data sample used for the estimation. If the model describes the data-generating process correctly, then the maximum likelihood estimates are asymptotically efficient and a single-plug-in model that uses the same parameter estimates for all forecasting iterations is preferred. However, in practice models are not correct. In-sample prediction errors at longer horizons can then contain systematic information not included in the one-step-ahead errors, and it is desirable to incorporate such information into the estimation procedure. The literature discussing such procedures starts probably with Cox (1961), who provides explicit formulas for predictors that exploit information from in-sample prediction errors at longer horizons for a stationary AR(1) process. Tiao and Xu (1993) extend such considerations to ARIMA processes. Xia and Tong (2011) denote the family of approaches to model fit other than one-step-ahead prediction errors as feature matching.[2] Recently, Schorfheide (2005) and Kapetanios et al. (2015) consider a vector of prediction errors for different horizons in estimation and forecasting in the VAR and DSGE model frameworks, respectively.

The approach to multi-step forecasting introduced in this paper follows this line of research and exploits information from in-sample prediction errors at longer horizons. The model coefficients are thus allowed to change in the direction of minimizing the m-step-ahead prediction errors. So, for the first forecasting iteration (the one-step-ahead forecast) the estimation method is basically maximum likelihood, except that some prior information on the model parameters is imposed. Other forecasting iterations then take into account both the estimation results from the previous forecasting iterations and the in-sample prediction errors for the respective horizon. The approach is reminiscent of the direct forecasting method, in which the estimation of a horizon-specific model is related to the prediction errors at the corresponding horizon. The presented procedure thus in a way represents a combination of iterated and direct forecasting. The weights of the forecasting techniques are determined according to how close the resulting models are to the true data-generating process.

The combination of direct and iterated forecasting could be an answer to the trade-off between estimation bias and estimation variance that is an inherent feature of comparisons of the two basic forecasting techniques (Findley, 1983). The bias resulting from possible

[2] There can be other motivations for using m-step-ahead prediction errors when fitting a model different from the forecasting performance of the model. For example, Tonner and Bruha (2014) use the eight-step-ahead prediction error reflecting the length of monetary policy horizon.

misspecification of the benchmark model multiplied by the iterating one-step-ahead forecasts can be corrected by taking into account the systematic relations between two distant data points beyond what can be inferred from the relations between adjacent data points. On the other hand, the inefficiency of direct forecasting can be diminished by the iterative nature of the procedure and by the fact that the iterated approach produces more efficient parameter estimates.[3]

The presented approach is close to Kapetanios et al. (2015). In contrast to that paper, the approach in this paper introduces a fully fledged Bayesian perspective and allows for different model coefficients for different forecasting horizons when iterating one-step-ahead forecasts. On the other hand, the possibility of changing coefficients in each iteration precludes joint estimation of all the parameters appearing in the forecasting procedure because of the size of the parameter vector. Therefore, the trade-off between the forecasting performances at two different horizons and their potential optimization across forecasting iterations is not a subject of the estimation procedure.

The forecasting performance of the proposed methodology is examined on a medium-scale vector autoregression that includes the same variables as the DSGE model of the US economy in Smets and Wouters (2007). More precisely, the data sample starts in 1959Q1 and the forecasting performance is examined on the data observed during the period 1998Q4–2016Q1. Both point and density forecasts are discussed. For point forecasts it turns out that adjusted iterated forecasts outperform both standard iterated and direct forecasts. The result is even more clear-cut for density forecasts. Adjusted iterated forecasts thus represent a technique for dealing with the trade-off between potentially biased iterated forecasting and inefficient direct forecasting.

The rest of the paper is organized as follows. Section 2 presents the model and the standard iterated forecasting procedure. Section 3 describes how the model coefficients are adjusted to minimize the m-step-ahead prediction errors and how the adjusted iterated forecasts are constructed. Section 4 presents the dataset and describes the specification of priors and the set-up of the forecasting performance exercise. Section 5 discusses the results and offers some robustness checks. Finally, Section 6 concludes and Appendix A provides a brief description of the data.

2. Model and iterated forecasting

To demonstrate the general principle for the adjustment of model coefficients in an iterated forecasting process, a standard VAR model is considered:

$$
\begin{aligned}
y_t &= C + B_1 y_{t-1} + \ldots + B_p y_{t-p} + \varepsilon_t \\
\varepsilon_t &\sim N(0, \Sigma),
\end{aligned}
\tag{1}
$$

[3] A detailed empirical comparison of direct and iterated forecasts can be found in Marcellino et al. (2006).

where y_t is an $n \times 1$ vector of endogenous variables, ε_t is an $n \times 1$ vector of exogenous shocks, C, $B_1,\ldots,B_p$ are an $n \times 1$ vector and $n \times n$ matrices of constants and AR parameters, respectively, and Σ is an $n \times n$ matrix of error covariances. Model (1) is the model for the vector of data $y^T \equiv \{y_1,\ldots,y_T\}$.

System (1) can be rewritten as follows:

$$\begin{aligned} y_t &= X_t\beta + \varepsilon_t \\ \varepsilon_t &\sim N(0,\Sigma), \end{aligned} \tag{2}$$

where $X_t \equiv I_n \otimes x_t'$, $x_t = [1, y_{t-1}',\ldots,y_{t-p}']$, and $\beta \equiv vec[C, B_1,\ldots,B_p]'$.

The one-step-ahead forecast (or predictive density) is then:

$$\begin{aligned} y_{T+1} &= X_{T+1}\beta + \varepsilon_{T+1} \\ \varepsilon_{T+1} &\sim N(0,\Sigma). \end{aligned} \tag{3}$$

Note that y_{T+1} is an $n \times 1$ vector of random variables with a distribution given by the distribution of the estimated model parameters β and Σ.

Iterating (3) forward, forecasts for other forecasting horizons, $h = 2,\ldots,H$, can be generated. For example, the two-step-ahead forecast is as follows:

$$\begin{aligned} y_{T+2} &= X_{T+2}\beta + \varepsilon_{T+2} \\ \varepsilon_{T+2} &\sim N(0,\Sigma) \end{aligned} \tag{4}$$

where the matrix $X_{T+2} = I_n \otimes [1, \hat{y}_{T+1}', y_T',\ldots,y_{T+1-p}']$ contains the observed data and the results from the first forecasting iteration. The iterated forecasts are usually based on the same estimated model parameters $\hat{\beta}$ and $\hat{\Sigma}$ for each forecasting iteration.

Model (2) is estimated employing the Normal inverse Wishart prior:

$$\begin{aligned} \beta \mid \Sigma &\sim N(b, \Sigma \otimes \Omega) \\ \Sigma &\sim iW(k, \Theta). \end{aligned} \tag{5}$$

The prior materializes the prior belief that the variables follow a process close to a random walk. Combining prior (5) with the likelihood function yields the posteriors of the model parameters. The likelihood function (conditional on the initial p observations) can be expressed as a product of the conditional probability densities:

$$p\left(y^T \mid \beta, \Sigma\right) = \prod_{t=p+1}^{T} p\left(y_t \mid y^{t-1}, \beta, \Sigma\right). \tag{6}$$

Equation (6) can be interpreted such that the likelihood is the value of the probability density in the case where the model produces a one-step-ahead in-sample prediction equal to the observed values. Equivalently, it represents the value of the probability density that the one-step-ahead prediction errors are zero. So, maximizing the likelihood means minimizing the in-sample one-step-ahead prediction errors. If the prior distributions are flat, the posteriors are proportional to the likelihood and the above-mentioned interpretation of likelihood also applies to the posteriors.

The Normal inverse Wishart prior belongs to the family of natural conjugate priors and thus the posteriors follow the same distributions. Moreover, the formulas for the posterior distributions of the model parameters can be expressed in closed form:

$$\begin{aligned} &\beta \mid \Sigma, y^T \sim N\left(\hat{\beta}, \Sigma \otimes \hat{\Omega}\right) \\ &\Sigma \mid y^T \sim iW\left(d, \hat{\Theta}\right) \end{aligned} \tag{7}$$

where $\hat{\Omega} = \left(x'x + \Omega^{-1}\right)^{-1}$ and $\hat{\beta} = \hat{\Omega}\left(x'y + \Omega^{-1}\mathrm{B}\right)$. Next, $x \equiv \left[x_{p+1}, \ldots, x_T\right]'$, $y \equiv \left[y_{p+1}, \ldots, y_T\right]'$, and B is constructed such that the columns are created from the prior coefficients for the parameters in each equation. Finally, the scale parameter of the posterior for the error covariance matrix equals $\hat{\Theta} = \Theta + SSR + \left(\hat{\mathrm{B}} - \mathrm{B}\right)' \Omega^{-1} \left(\hat{\mathrm{B}} - \mathrm{B}\right)$, where the term SSR denotes the sum of the squared residuals from the regression with the posterior of the AR coefficients. The degrees of freedom parameter $d = k + T - p$.

The full distribution of the one-step-ahead forecast y_{T+1} is matricvariate-t:

$$y_{T+1} \mid y^T \sim MT\left(X_{T+1}\hat{\beta}, \left(X_{T+1}\left(\Sigma \otimes \hat{\Omega}\right)X'_{T+1} + 1\right), k + T - p, \hat{\Theta}\right). \tag{8}$$

The distributions of the iterated forecasts for other forecasting horizons $h = 2, \ldots, H$ do not have an analytical form and are simulated.

3. Adjusted iterated forecasting

In this section, we show how the forecasting iteration process is adjusted to take into account the in-sample prediction errors for higher forecasting horizons. In the first forecasting iteration, i.e., for the forecasts for one period ahead ($h = 1$), we stick to the original formula (3) and the original parameter estimates, which are now indexed by the forecasting iteration, i.e., $\hat{\beta}^{(1)}$ and $\hat{\Sigma}^{(1)}$. Such estimates take into account the one-step-ahead prediction performance within the data sample, as the likelihood function reflecting the probability of zero forecasting errors is combined with the prior distributions in (5).

Analogously, we can discuss the two-step-ahead prediction errors for the observed data given the estimation results from the first forecasting iteration. More precisely, instead of the probability density of the one-step-ahead forecast $p\left(y_t \mid y^{t-1}, \beta^{(1)}, \Sigma^{(1)}\right)$ our focus moves to the probability of zero two-step-ahead forecast errors in period t:

$$p\left(y_t \mid y^{t-2}, \hat{\beta}^{(1)}, \hat{\Sigma}^{(1)}, \beta^{(2)}, \Sigma^{(2)}\right), \tag{9}$$

which can be reformulated using the fitted values from the first forecasting iteration $\hat{y}^{(1)}_{t-1} = X_{t-1}\hat{\beta}^{(1)}$ as follows:

$$p\left(y_t \mid y^{t-2}, \hat{\beta}^{(1)}, \hat{\Sigma}^{(1)}, \beta^{(2)}, \Sigma^{(2)}\right) = p\left(y_t \mid y^{t-2}, \hat{y}^{(1)}_{t-1}, \beta^{(2)}, \Sigma^{(2)}\right). \tag{10}$$

Put differently, the likelihood of the observed data given the model parameters for the model of two-step-ahead iterated forecasts:

$$\begin{aligned} & y_t = C^{(2)} + B^{(2)}_1 \hat{y}^{(1)}_{t-1} + B^{(2)}_2 y_{t-2} + \ldots + B^{(2)}_p y_{t-p} + \varepsilon_t \\ & \varepsilon_t \sim N\left(0, \Sigma^{(2)}\right) \end{aligned} \tag{11}$$

can be expressed as follows:

$$\begin{aligned} & p\left(y^T \mid \beta^{(2)}, \Sigma^{(2)}, \hat{y}^{(1)}_{T-1}, \ldots, \hat{y}^{(1)}_{p+1}\right) = \prod_{t=p+1}^{T} p\left(y_t \mid y^{t-2}, \beta^{(2)}, \Sigma^{(2)}, \hat{y}^{(1)}_{t-1}\right) = \\ & = \prod_{t=p+1}^{T} p\left(y_t \mid y^{t-2}, \hat{\beta}^{(1)}, \hat{\Sigma}^{(1)}, \beta^{(2)}, \Sigma^{(2)}\right) \end{aligned} \tag{12}$$

Formula (12) suggests that the likelihood of model (11) expresses the probability density of zero two-step-ahead forecast errors conditional on the first forecasting iteration. The likelihood can be combined with the prior on the model parameters, yielding the posterior of the model parameters for the second forecasting iteration: $\beta^{(2)}$ and $\Sigma^{(2)}$.

When estimating model (11) the important point is that fitted values $\hat{y}^{(1)}_{t-1}$ need to be treated as a random variable. This fact affects both the specification of the priors for the second forecasting iteration and the Bayesian inference itself.

For the second forecasting iteration we retain our prior belief that the process $\{y_t\}$ follows a random walk in the form of the Normal inverse Wishart prior. In addition, the uncertainty related to the fitted values needs to be accounted for. Given the estimation results from the first forecasting iteration, the fitted values used in the second iteration are distributed normally with the following moments:

$$E\left[\hat{y}_{t-1}^{(1)}\right] = X_{t-1}\hat{\beta}^{(1)}$$
$$\text{var}\left[\hat{y}_{t-1}^{(1)}\right] = X_{t-1}\,\text{var}\left[\hat{\beta}^{(1)}\right]X'_{t-1}, \tag{13}$$

where $t = p+2,...,T+1$.

The fitted value $\hat{y}_{t-1}^{(1)}$ includes estimation uncertainty related to $\hat{\beta}^{(1)}$ but not uncertainty related to the shock realized at time *t-1*. The fitted value $\hat{y}_{t-1}^{(1)}$ and the observed value y_{t-1} differ in the realized shock ε_{t-1} for which we have formulated a prior in the first forecasting iteration. The shock becomes a part of the disturbance term at time *t* in the model for the second forecasting iteration (11). The sum of two i.i.d. normally distributed variables has the same mean and double the variance. Therefore, consistency of the priors on the error covariance implies doubled prior error covariance for the second forecasting iteration.

Next, given that the prior from the first forecasting iteration holds exactly, the distribution of the fitted value $\hat{y}_{t-1}^{(1)}$ is centered on y_{t-2} with a variance proportional to the prior variance of the AR parameters. Model (11) is then 'close' to the model $y_t = C + B_1^{(2)}y_{t-2} + B_2^{(2)}y_{t-2} + B_3^{(2)}y_{t-3}... + B_p^{(2)}y_{t-p} + \varepsilon_t$. The random walk prior belief implies that $B_1^{(2)}$ should be centered on unity, because if $\{y_t\}$ follows a random walk, then it holds that

$$y_t = y_{t-1} + \varepsilon_t^{RW} = y_{t-2} + \varepsilon_{t-1}^{RW} + \varepsilon_t^{RW}. \tag{14}$$

The prior variance for the AR parameter $B_1^{(2)}$ should be lowered in comparison to the prior variance of the coefficient from the first forecasting iteration $B_1^{(1)}$ because the presence of random variable $\hat{y}_{t-1}^{(1)}$ already imposes a degree of uncertainty that $B_1^{(2)}$ is centered on unity. So, the tightness of the prior variance of the AR parameter at the first lag is half of the assumed prior variance from the first forecasting iteration. This choice is discussed in detail in the section describing prior specifications.

Finally, the Kronecker structure of the prior variance on the AR terms in the Normal inverse Wishart distribution implies that multiplying the scale of the prior on the error covariance matrix proportionally affects the prior on the variance on the AR parameters. Such multiplication is compensated for by multiplying the overall tightness by $\left(1/\sqrt{2}\right)$.

The presence of the random variable in the set of RHS variables in model (11) also affects the Bayesian inference. Analytical formulas are not available and the marginal posteriors of the model parameters for the second forecasting iteration are simulated using MC sampling with the following steps:

1) Initialize the values of $\beta^{(2)}$ and $\Sigma^{(2)}$.
2) Given the sample from the posterior distributions of $\beta^{(1)}$ estimated in the first forecasting iteration, take a random draw $\hat{y}_{t-1}^{(1)}$ according to (13) for $t = p+2,...,T+1$.
3) Given the observed data y^T and the draw $\hat{y}_{t-1}^{(1)}$, take a random draw of $\beta^{(2)}, \Sigma^{(2)}$ following the standard formulas for the Normal inverse Wishart conjugate priors.[4]
4) Repeat steps 2 and 3 many times and take summary statistics of the draws of the model parameter subsets.

Iterated forecasts for horizon $h = 2$ are simulated by using the draws of $\beta^{(1)}$ and $\Sigma^{(1)}$ from the first forecasting iteration and then by taking draws $\beta^{(2)}$ and $\Sigma^{(2)}$ from the second forecasting iteration.

For other forecasting iterations, the procedures for the adjustment of model coefficients and the simulation of forecasts are analogous to the case $h = 2$.

4. Data, priors, and set-up of forecasting exercise

The data set includes real GDP (RGDP), the GDP deflator (PGDP), consumption (Cons), investment (GDPInv), hours worked (Emp. Hours), wages (Real Comp/Hour), and the federal funds rate (FedFunds). The variables are of quarterly frequency covering the period 1959Q1–2016Q1.[5] A list of variables can be found in Appendix A. All variables are in annualized log levels except for the federal funds rate, which is in levels divided by 100. Figure 1 presents the data set. It can be seen that both trending variables and time series that are presumably stationary are included. Such diversity can indicate whether or not the effect of the proposed methodology on forecasting performance is dependent on the basic properties of the time series.

[4] The implementation is such that the means of the posterior distributions for $\beta^{(2)}$ and $\Sigma^{(2)}$ are taken. Knowledge of the analytical form of the posterior distributions is thus exploited. As a robustness check, ten random draws of $\beta^{(2)}, \Sigma^{(2)}$ are taken instead of a value equal to the posterior mean. This change results in slightly more imprecise estimates of $\beta^{(2)}$ and $\Sigma^{(2)}$.

[5] Source: Federal Reserve Bank of St. Louis Database (FRED).

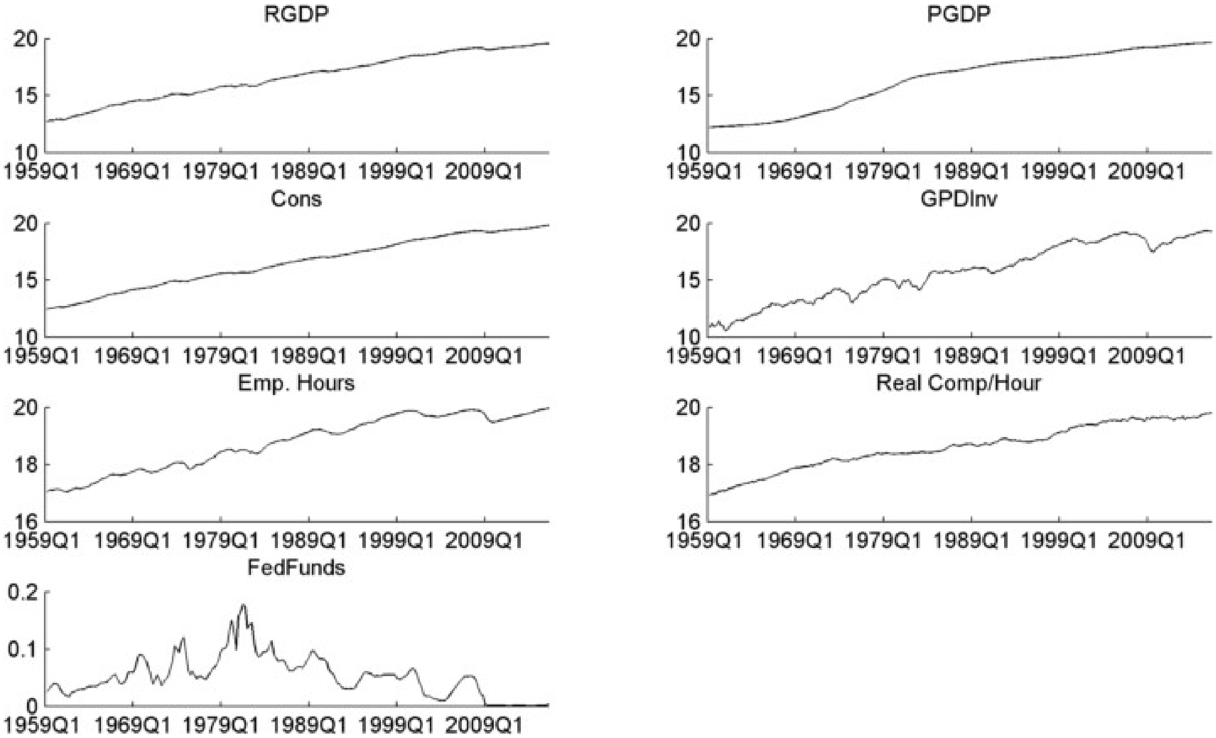

Figure 1: Endogenous variables entering the model estimation.

The parameters of the inverse Wishart prior assumed for the error covariance matrix Σ are set so that the degrees of freedom parameter $d = n + 2$, which is the minimum value that guarantees the existence of a mean of the distribution. The scale matrix, Θ, is a diagonal matrix with the estimated error variances of the AR(1) regressions of the respective LHS variable on its first own lag.

The specification of the prior distributions of the AR parameters β is such that in all equations the mean of the coefficient on the first own lag of the LHS variable is one and that for the rest of the coefficients is zero. The prior mean on the intercepts is zero as well. The variance of the prior distribution for the AR parameters conditional on Σ is such that its elements are defined in the following way: the coefficient in the *i*-th equation for the *s*-th lag of the *j*-th variable is the following:

$$\frac{\lambda^2}{s^2} \frac{\Sigma_{ii}}{\Theta_{jj}\left(d - n - 1\right)}, \tag{15}$$

where parameter λ represents the overall tightness of the prior variance. The prior variance on the coefficient at the intercept is 10^4.

As discussed above, the priors change with the forecasting iteration h. First, the scale matrix from the inverse Wishart prior is multiplied by $\min(h, p)$ to account for the inclusion of fitted values in the models for higher forecasting iterations. To filter out the effect of such rescaling on the prior variance of the AR parameters, the overall tightness λ is multiplied by $\sqrt{\min(h, p)}$. Finally, to account for the uncertainty imposed by the uncertain fitted values in the regression, the overall tightness of the prior variance of the AR parameters at the fitted values is divided by 2. The value of the tightness parameter for the first forecasting iteration is set equal to 0.2, which is a standard value in the literature. The prior variance on the intercept does not change across forecasting iterations.

The intuition behind tightening the prior on the AR parameters at fitted values is captured in Figure 2. In the second forecasting iteration two distributions are combined in the form of their product—the distribution of fitted values from the first forecasting iteration ($\hat{y}_{t-1}^{(1)}$) and the prior on the respective parameter at the fitted value ($B_1^{(2)}$). The product cannot be expressed analytically and Figure 2, panel c, shows its simulation. If one compares the simulated product and prior imposed in the first forecasting iteration ($B_1^{(1)}$) they are very similar. And this is the purpose of prior tightening—to ensure that the prior uncertainty related to independent variables is similar in all forecasting iterations (because we try to model the same dependent variable in all forecasting iterations).

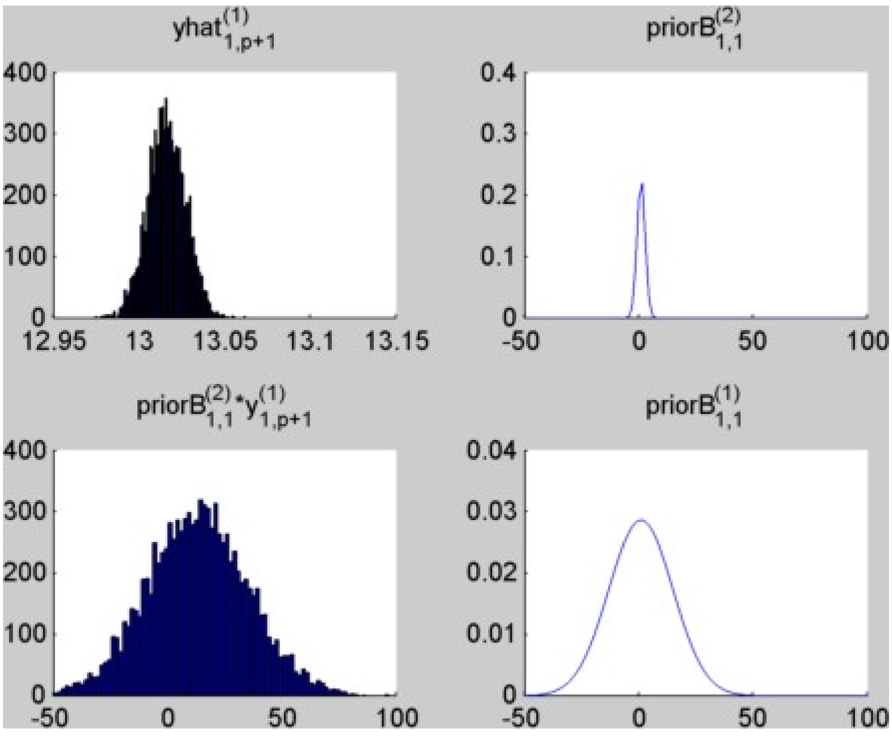

Figure 2. Prior tightening in the second forecasting iteration.
Note: The simulated fitted value of real GDP from the first forecasting iteration (panel a), the prior on the coefficient at the first lag of real GDP in the equation for real GDP in the second forecasting iteration (panel b), and the simulated product of the two (panel c). Panel d shows the prior on the coefficient at the first lag of real GDP in the equation for real GDP in the first forecasting iteration. The estimation is done on the full sample.

Following Giannone et al. (2015), the number of lags in (1) is set to five. The sampler contains 5,000 iterations. Convergence is tested using standard measures: the autocorrelation of parameter draws produced by the sampler, the inefficiency factor, and the measure of the number of draws needed to get a stationary distribution from the sampler (Raftery and Lewis, 1992). All measures suggest convergence of the sampler. The results are available upon request.

The pseudo-out-of-sample forecasting exercise is based on 70 observations between 1998Q4 and 2016Q1. So, in the first round the models are estimated on the period 1959Q1–1998Q3 and forecasts for up to 12 quarters ahead are simulated. The iterated forecasts, the adjusted iterated forecasts, and the direct forecasts are then compared in terms of point and density forecasting accuracy. The second round then uses data covering the period 1959Q1–1998Q4, etc.

The point forecasting accuracy is computed using the standard mean squared forecast error (MSFE). The Diebold-Mariano test of equal forecasting accuracy is carried out, correcting for autocorrelation of the residuals. The density forecasting accuracy is assessed using the Kullback-Leibler Information Criterion. Minimization of the criterion can be rewritten as maximization of the expected logarithmic score, which is estimated by the average logarithmic score:

$$\frac{1}{N}\sum_{t\in A}\ln f_{t+h,t}\left(\bar{y}_{i,t+h}\right), \tag{16}$$

where $\bar{y}_{i,t+h}$ is the ex-post realization of the variable and $f_{t+h,t}$ is the simulated posterior density of that variable computed at time t at forecasting horizon h.

5. Results

Table 1 demonstrates how the in-sample fit of models that take into account prediction errors for longer forecasting horizons improves. The in-sample fit is measured as the mean squared error of the fitted values constructed for a particular horizon. The fits of models with coefficients adjusted for higher-period prediction errors and the standard model are compared. The positive values in the table suggest that the in-sample fit of the model with adjusted coefficients is superior for all horizons (exhibiting lower mean square errors). This is not surprising, as simply a higher number of parameters is used to explain the observed data. However, the increase in fit demonstrates that the original model is not correctly specified. If the model for the first forecasting period described the data generating process correctly, the improvement in data fit would not be observed.

Table 1: The mean difference of the in-sample fit of models used for iterated and adjusted iterated forecasting.

	RGDP	PGDP	Cons	GDPInv	Emp. Hours	Real Comp /Hour	FedFunds
Horizon:							
1	0	0	0	0	0	0	0
2	0.0001	0.0000	0.0000	0.0005	0.0001	0.0000	0.0000
3	0.0000	0.0000	0.0000	0.0004	0.0000	0.0000	0.0000
4	0.0001	0.0000	0.0001	0.0012	0.0001	0.0001	0.0000
5	0.0002	0.0000	0.0001	0.0028	0.0002	0.0001	0.0000
6	0.0003	0.0001	0.0002	0.0051	0.0003	0.0002	0.0000
7	0.0007	0.0003	0.0004	0.0096	0.0007	0.0003	0.0000
8	0.0012	0.0007	0.0008	0.0168	0.0014	0.0006	0.0001
9	0.0023	0.0017	0.0015	0.0290	0.0025	0.0010	0.0001
10	0.0042	0.0037	0.0029	0.0502	0.0044	0.0018	0.0002
11	0.0077	0.0081	0.0056	0.0886	0.0078	0.0033	0.0004
12	0.0144	0.0173	0.0105	0.1596	0.0139	0.0059	0.0007

Notes: The in-sample fit is estimated using the squared differences between the fitted and observed values. The models are estimated on the full sample.

Next, the point forecasting performance is examined. Figure 3 reports the MSFEs of the forecasts produced by standard iterated forecasting (black solid line) and by adjusted iterated forecasting (red dashed line). The model which allows for changes in coefficients exhibits lower MSFEs for almost all variables and horizons. The MSFEs are expressed in units of the respective variable, so they are not directly comparable across variables. However, Figure 3 suggests that for some variables, the adjusted forecasting procedure can lower the MSFE to half of the MSFE of iterated forecasts. The differences in MSFEs between the two approaches are reported in Table 2. The table also contains the results of the Diebold-Mariano test of equal forecasting accuracy of the two forecasting techniques.

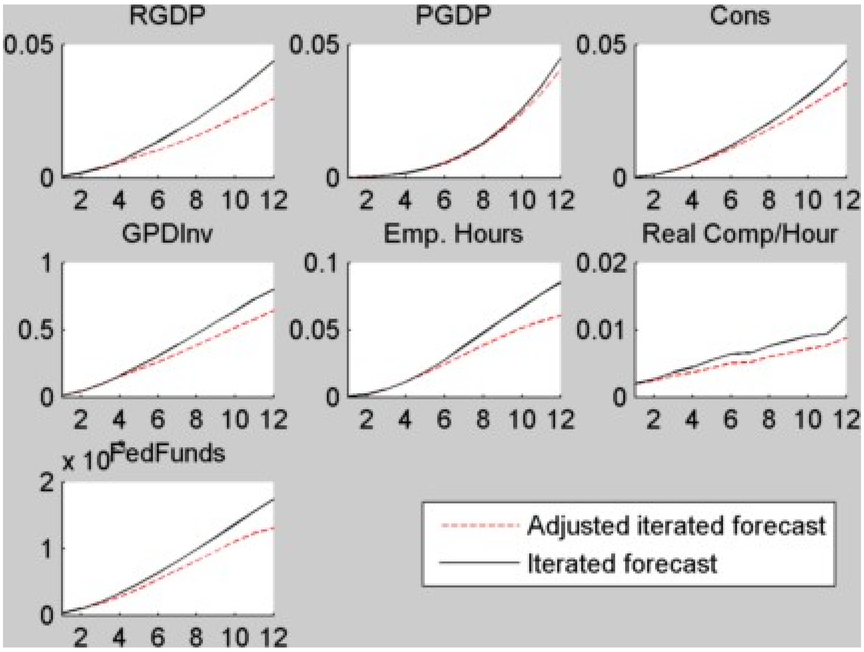

Figure 3. Mean square forecast errors at a particular forecasting horizon for the adjusted iterated forecasts and iterated forecasts.

Table 2. The difference between the MSFEs of iterated and adjusted iterated forecasts.

Horizon:	RGDP	PGDP	Cons	GDPInv	Emp. Hours	Real Comp /Hour	FedFunds
1	0	0	0	0	0	0	0
2	0.0001	0.0000	0.0000	-0.0024	-0.0001*	0.0002*	0.0000
3	0.0002	0.0000	0.0000	-0.0039	-0.0003	0.0005***	0.0000
4	0.0006	0.0000	0.0002	0.0024	-0.0001	0.0008***	0.0000*
5	0.0017*	0.0000	0.0006	0.0209*	0.0010*	0.0012***	0.0001**
6	0.0032**	0.0003	0.0010	0.0437**	0.0029***	0.0013*	0.0001***
7	0.0048**	0.0006	0.0017	0.0656**	0.0061***	0.0014	0.0001***
8	0.0061*	0.0003	0.0024	0.0829***	0.0089***	0.0016	0.0002***
9	0.0079**	0.0008	0.0034	0.1011***	0.0122***	0.0018	0.0002***
10	0.0093*	0.0014	0.0044	0.1216***	0.0152***	0.0020	0.0003**
11	0.0120**	0.0027	0.0059	0.1496***	0.0202***	0.0017	0.0003**
12	0.0141*	0.0041	0.0086	0.1623***	0.0245***	0.0030	0.0004**

Note: *, **, and *** denote significance at the 10%, 5%, and 1% levels of confidence for the Diebold-Mariano test of equal forecasting accuracy.

Table 2 shows that all except one of the statistically significant differences in forecasting performance are observed only for the case where adjusted iterated forecasting is more accurate than standard iterated forecasting. Moreover, the magnitude of the difference is substantial and in some cases is close to the average difference in the variable between two adjacent periods. Finally, the improvement in forecasting accuracy is observed regardless of whether the variables exhibit trends or not.

Turning to the density forecasting performance, Table 3 suggests that adjusting the coefficients improves the accuracy of the density forecasts in all cases (i.e., the average logarithmic score for adjusted iterated density forecasts is higher than that for standard iterated density forecasts). While the median iterated forecasts are in some cases comparable to the median adjusted forecasts, the comparison of whole densities suggests a clear preference for adjusting model coefficients in iterated forecasting.

Table 3. The difference between the average logarithmic scores of adjusted iterated and iterated forecasts.

Horizon:	RGDP	PGDP	Cons	GDPInv	Emp. Hours	Real Comp /Hour	FedFunds
1	0	0	0	0	0	0	0
2	0.08	0.12	0.07	0.02	0.07	0.20	0.00
3	0.12	0.18	0.18	0.10	0.15	0.31	0.03
4	0.24	0.23	0.28	0.23	0.31	0.36	0.07
5	0.37	0.29	0.39	0.37	0.49	0.41	0.13
6	0.44	0.37	0.45	0.46	0.65	0.41	0.22
7	0.52	0.44	0.53	0.56	0.82	0.42	0.28
8	0.57	0.49	0.61	0.66	0.94	0.43	0.33
9	0.66	0.60	0.68	0.75	1.09	0.47	0.41
10	0.73	0.71	0.77	0.91	1.24	0.52	0.49
11	0.84	0.81	0.85	1.06	1.41	0.55	0.55
12	0.92	0.93	0.93	1.21	1.59	0.64	0.62

In producing adjusted iterated multi-step forecasts, a model with different coefficients is used in each iteration. As an example, Figure 4 shows the evolution of the estimates of the selected AR parameters in the equation for real GDP. It reports the evolution of the intercept (C_1), the coefficient on the first lag of real GDP (B_{11}), and the coefficients on the first lags of the GDP deflator and consumption (B_{12} and B_{13}). From the second forecasting iteration, the coefficient on the own lag of real GDP moves close to unity from its original value (denoted by the red dashed line). Similarly, after three forecasting iterations, the coefficients on the first lag of the other reported variables are close to zero. The value of the intercept converges to a positive figure. Not surprisingly, it turns out that the best prediction at longer horizons is obtained by taking the previous period fitted value and adding the mean of real growth, which is estimated by the intercept. Note that real GDP enters the model in log-level form. The purpose of the estimation procedure is to choose the most accurate way of moving from the information included in the observed variables when forecasting short horizons to longer horizons, where the estimated long-run value dominates. Put differently, maximum likelihood represents a high-pass filter, whereas the procedure for adjusting the coefficients for longer forecasting horizons represents a low-pass filter.

A slightly different picture can be seen when one looks at the evolution of the coefficients in the equation for the federal funds (FF) rate—see Figure 5. The interest rate enters the analysis in levels, but does not exhibit clear trending behavior like the rest of the variables. The intercept approaches zero and the evolution of the parameters at their own lag is such that their tilting enables the model to accelerate towards the sample mean of the FF rate.

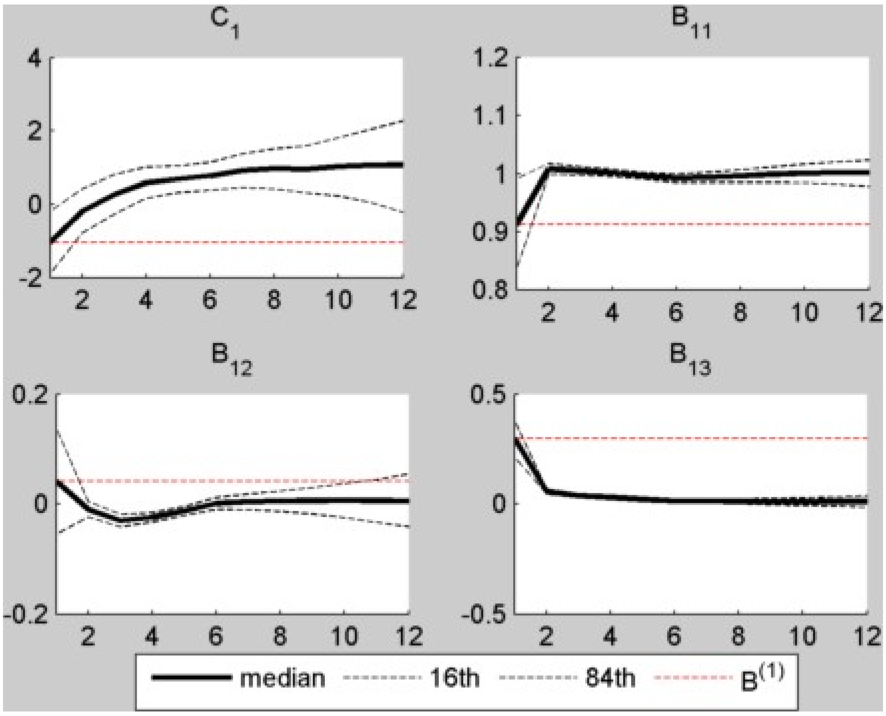

Figure 4: Evolution of the coefficients in the equation for real GDP.

Note: Panels indicate evolution of the intercept (C_1), the first lag of real GDP (B_{11}), the first lag of the GDP deflator (B_{12}), and consumption (B_{13}). The red dashed line indicates the median estimate for the first forecasting iteration.

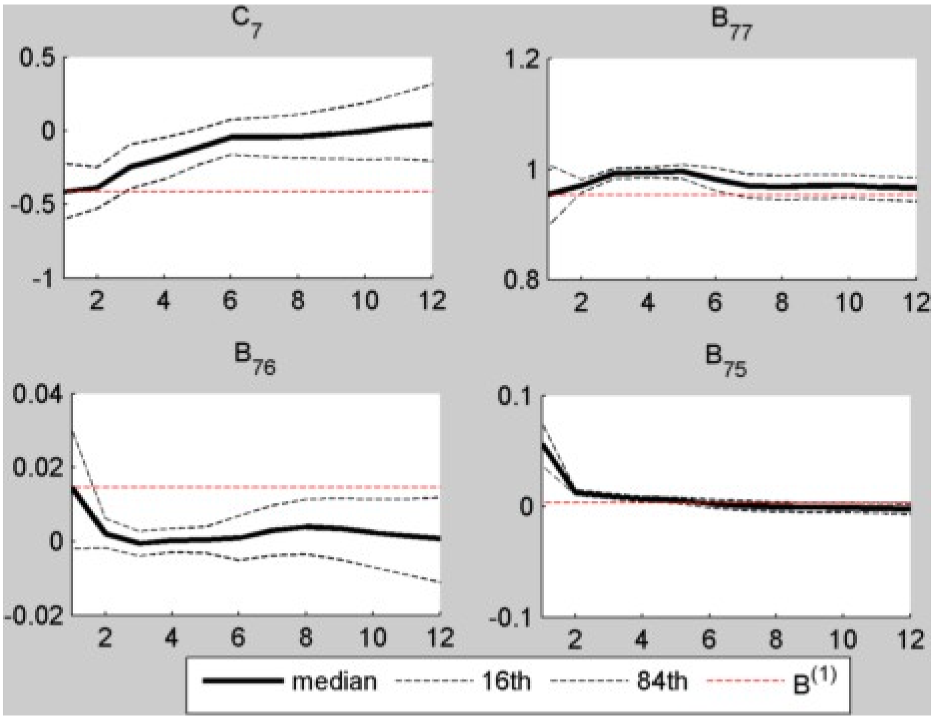

Figure 5: Evolution of the coefficients in the equation for the federal funds rate.

Note: Panels indicate evolution of the intercept (C_7), the first lag of the federal funds rate (B_{77}), the first lag of real wages (B_{76}), and hours worked (B_{75}). The red dashed line indicates the median estimate for the first forecasting iteration. The estimation is done on the full sample.

Finally, Figure 6 shows the evolution of selected coefficients of the error covariance matrix over the forecasting iterations. The presented diagonal elements increase, reflecting the increasing prior on the error variances. As discussed above, the prior is rescaled to reflect the fact that higher forecasting horizons contain higher uncertainty stemming from shocks.

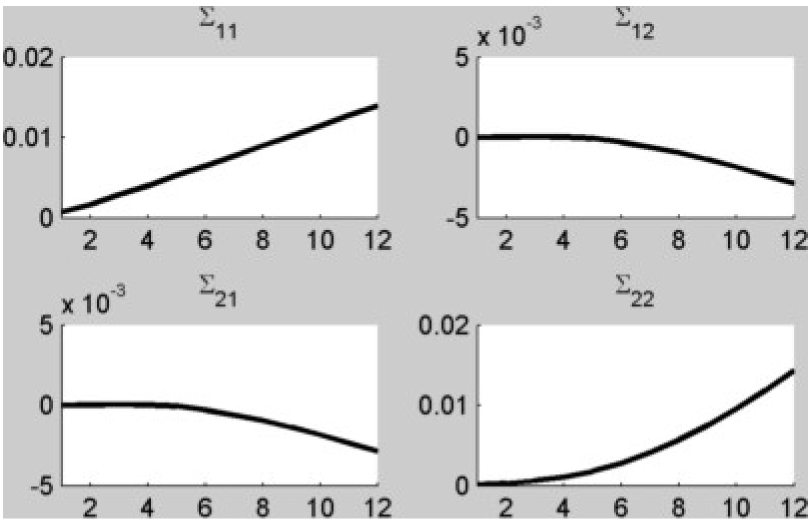

Figure 6: Evolution of selected elements of the error covariance matrix.

Note: Panels indicate evolution of the median of the error variance in the real GDP equation (Σ_{11}) and the federal funds equation (Σ_{22}) and the covariances between the two. The estimation is done on the full sample.

5.1 Robustness issues

The specification of priors follows standard values from the literature. The only exception is the parameter λ , representing the overall tightness. For the first forecasting iteration it equals 0.2, which is a standard value. However, for other forecasting iterations no standard values are available. The only clue follows from the fact that the fitted values used in the other forecasting iterations are uncertain and the overall tightness should reflect this fact by forcing the priors towards their prior means. A robustness check regarding the overall tightness is, however, necessary.

The robustness exercise consists of two extremes. The first exercise assumes that the overall tightness ignores the uncertainty in the fitted values in the sense that it does not decrease for coefficients at fitted values and the prior on the error variance is not re-scaled by 2. On the other hand, the second exercise assumes a more profound drop in the overall tightness of the priors for parameters at fitted values. The overall tightness is not ½ of that from the first forecasting iteration, but 1/3. The specific values of the overall tightness for parameters at fitted values for different forecasting iterations are reported in Table 4.

Table 4. Alternative values of the overall tightness of the prior on AR parameters at fitted values.

horizon:	1	2	3	4	5
Initial calibration	0.2	0.0707	0.0577	0.0500	0.0447
Loose lambda	0.2	0.2000	0.2000	0.2000	0.2000
Tight lambda	0.2	0.0471	0.0385	0.0333	0.0298

Figure 7 shows MSFEs for the standard iterated forecasts and adjusted iterated forecasts with different overall tightness λ for the prior on parameters at fitted values. Ignoring the uncertainty related to fitted values often leads to worse forecasting performance in comparison to both standard iterated forecasts and adjusted iterated forecasts with the initial calibration of λ . On the other hand, more intensive tightening of prior variances results in very similar results to the initial calibration of λ .

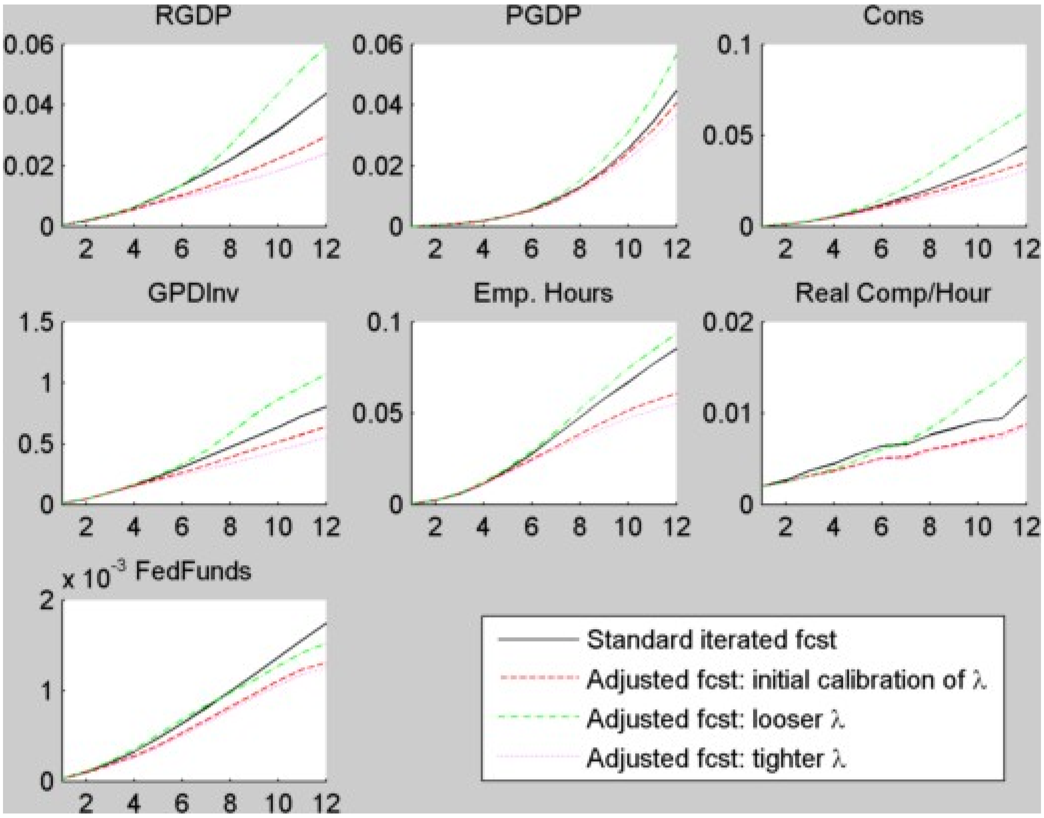

Figure 7. Mean square forecast errors of iterated forecasts and adjusted forecasts with different values of the overall tightness parameter.

Finally, restricting the data set for the evaluation of forecasting accuracy to the period 1998Q4–2016Q1 to examine the role of the Great Recession suggests that the Great Recession does not affect the conclusions about the superior forecasting performance of adjusted iterated forecasts. The detailed results are available upon request.

5.1 Adjusted iterated, iterated, and direct forecasting

This subsection focuses on comparing iterated forecasts, iterated forecasts adjusted according to the m-step-ahead in-sample prediction error performance, and direct forecasts. For direct forecasts, the same prior as for standard iterated forecasts is assumed. Figure 8 compares the MSFEs of the three forecasting techniques. Regarding the performance of iterated and direct forecasts, it turns out that no clear-cut conclusion can be drawn. The GDP deflator is forecasted more accurately by iterated forecasts for all horizons, but for other variables direct forecasts often outperform iterated forecasts. Interestingly, for short horizons of up to two quarters, the majority of the variables are better forecasted using iterated forecasts. Long horizons are almost exclusively better forecasted using the direct forecasting technique (the only exception being the GDP deflator). This result suggests the presence of bias, which is multiplied by iterating forecasts in the iterated forecasting technique.

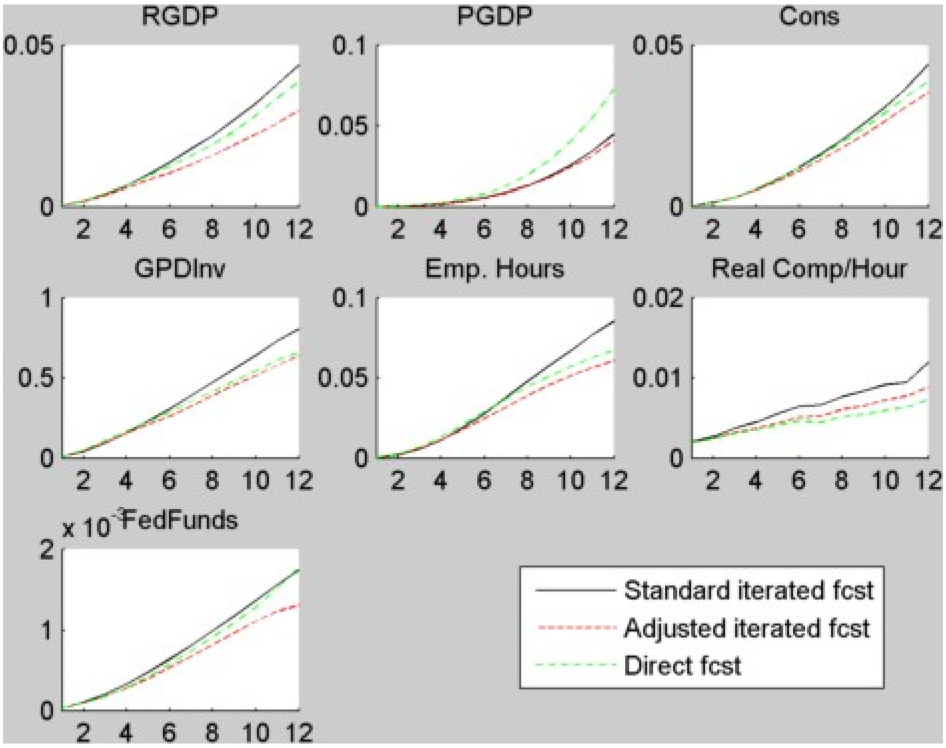

Figure 8: Mean square forecast errors at a particular forecasting horizon for the adjusted iterated forecasts, iterated forecasts, and direct forecasts.

Focusing on the comparison of adjusted iterated forecasts and direct forecasts, it can be concluded that in the vast majority of statistically significant cases, adjusted iterated forecasts outperform direct forecasts—see Table 5.

Table 5. The difference between the MSFEs of direct and adjusted iterated forecasts.

Horizon:	RGDP	PGDP	Cons	GDPInv	Emp. Hours	Real Comp /Hour	FedFunds
1	0	0	0	0	0	0	0
2	0.0001	0.0001**	0.0000	0.0042	0.0002**	0.0000	0.0000
3	0.0005*	0.0003**	0.0002	0.0105	0.0009**	-0.0001**	0.0000
4	0.0009	0.0006**	0.0005	0.0150	0.0019*	-0.0001*	0.0000
5	0.0013*	0.0013**	0.0007	0.0186	0.0029*	-0.0003	0.0000
6	0.0019	0.0027**	0.0010	0.0244	0.0041	-0.0005	0.0000
7	0.0027	0.0044**	0.0013	0.0304	0.0052	-0.0007	0.0001
8	0.0034	0.0069**	0.0017	0.0271	0.0057	-0.0009	0.0001
9	0.0043	0.0107**	0.0021	0.0256	0.0055	-0.0011	0.0001
10	0.0057	0.0161**	0.0026	0.0273	0.0055	-0.0013	0.0002
11	0.0078	0.0230***	0.0032	0.0295	0.0065	-0.0014*	0.0003
12	0.0089**	0.0320***	0.0034	0.0191	0.0066	-0.0016**	0.0005

Note: *, **, and *** denote significance at the 10%, 5%, and 1% levels of confidence.

So, adjusted iterated forecasts are superior in almost all cases to iterated forecasts and in the majority of cases also to direct forecasts. A similar conclusion can be drawn for density forecasts—see Table 6. Adjusted iterated density forecasts are more accurate than direct density forecasts in all cases. The accuracy of the density forecasts suggests efficiency of the model parameter estimates.

Table 6. The difference between the average logarithmic scores of direct and adjusted iterated forecasts.

Horizon:	RGDP	PGDP	Cons	GDPInv	Emp. Hours	Real Comp /Hour	FedFunds
1	0	0	0	0	0	0	0
2	-0.12	-0.10	-0.10	-0.10	-0.12	-0.15	-0.07
3	-0.23	-0.22	-0.23	-0.22	-0.29	-0.22	-0.19
4	-0.36	-0.37	-0.31	-0.33	-0.47	-0.27	-0.33
5	-0.48	-0.52	-0.42	-0.46	-0.66	-0.25	-0.46
6	-0.58	-0.71	-0.50	-0.53	-0.84	-0.21	-0.61
7	-0.70	-0.90	-0.55	-0.62	-1.03	-0.18	-0.77
8	-0.81	-1.12	-0.61	-0.73	-1.14	-0.16	-0.95
9	-0.93	-1.45	-0.70	-0.80	-1.22	-0.13	-1.16
10	-0.99	-1.73	-0.78	-0.87	-1.33	-0.10	-1.33
11	-1.09	-2.06	-0.88	-0.94	-1.47	-0.09	-1.53
12	-1.16	-2.40	-0.93	-1.03	-1.61	-0.09	-1.75

Figure 9 reports the evolution of selected estimated AR parameters in the equation for the FF rate for direct forecasting. When we compare it to Figure 5, it turns out that the model for adjusted iterated forecasting produces more efficient parameter estimates (as measured by the distance between the 16^{th} and 84^{th} percentiles of the posterior distribution of selected parameters). Adjusted iterated forecasts thus seem to enjoy the advantage of efficiency of parameter estimates in comparison to direct forecasts. Furthermore, adjusted iterated forecasts also share the advantage of direct forecasts in terms of forecasting robustness.

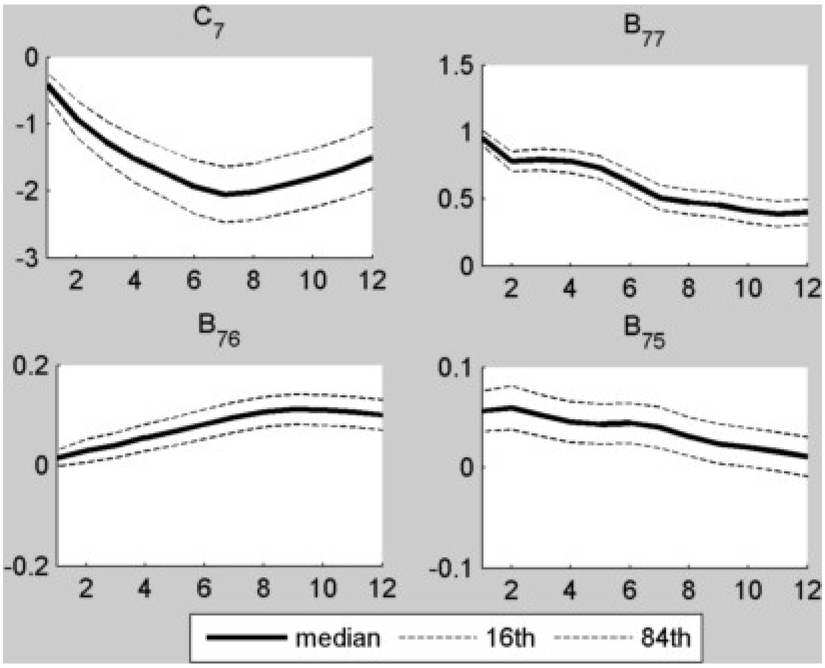

Figure 9. Evolution of the coefficients in direct forecasting in the equation for the federal funds rate

Note: Panels indicate evolution of the intercept (C_7), the first lag of the federal funds rate (B_{77}), the first lag of real wages (B_{76}), and hours worked (B_{75}). The estimation is done on the full sample.

6. Conclusions

The paper demonstrates how to adjust traditional iterated multi-step forecasts to get more accurate point and density forecasts. The adjustment draws on in-sample prediction errors for higher forecasting horizons. So, the approach extends forecasting based on one-step-ahead in-sample prediction errors. The point and density forecasting accuracy is demonstrated on a standard VAR model mimicking the Smets and Wouters (2007) DSGE model.

The model employed is a medium-scale VAR. For small-scale VARs the problem of misspecification is more profound and the improvement in accuracy would probably be greater. Similarly, DSGE models impose cross-coefficient restrictions, resulting in possible misspecification. So, the gain in forecasting performance discussed in this paper would presumably be higher for DSGE models than it is for their VAR counterparts.

The suggested approach can be viewed as a combination of iterated and direct forecasting. Iterated forecasting is represented by using one-step-ahead forecasts to get multi-step forecasts. Direct forecasting is present in taking into account the in-sample prediction error at a particular forecasting horizon. The combination of the two techniques is viable, as it could represent a response to the famous trade-off between bias and efficiency involved in the theoretical comparison of the two forecasting techniques. The results in the paper suggest that this is so. First, adjusted iterated forecasts are more precise in terms of mean squared forecasting errors. The bias is therefore lower. In addition, the density forecasting performance exercise suggests that adjusted iterated forecasting produces more accurate density forecasts, i.e., forecasts that are closer to the true densities of macroeconomic variables. Inefficient model parameter estimates would lead to less accurate density forecasts and thus the efficiency of iterated forecasting seems to carry over to adjusted iterated forecasting.

References

Cox, D. R. (1961): "Prediction by Exponentially Weighted Moving Averages and Related Methods", *Journal of the Royal Statistical Society: Series B*, 23, 414-422.

Findley, D. (1983): "On the Use of Multiple Models of Multi-Period Forecasting", *Proceedings of the Business and Statistics Section, American Statistical Association*, 528-531.

Giannone, D., Lenza, M., and G. E. Primiceri (2015): "Prior Selection for Vector Autoregression", *Review of Economics and Statistics,* 97(2), 436-451.

Kapetanios G., Price, S., and K. Theodoridis (2015): "A New Approach to Multi-Step Forecasting Using Dynamic Stochastic General Equilibrium Models", *Economics Letters*, 136(C), 237-342.

Marcellino, M., Stock, J. H., and M. W. Watson (2006): "A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series", *Journal of Econometrics*, 135, 499-526.

Raftery, A. E., and S. Lewis (1992): "How Many Iterations in the Gibbs Sampler?" In J. Bernardo, J. Berger, A. P. Dawid, and A. F. M. Smith (eds.), Bayesian Statistics, 763–773, Oxford University Press

Schorfheide, F. (2005): "VAR Forecasting Under Misspecification", *Journal of Econometrics*, 128, 99-136.

Smets, F., and R. Wouters (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach", *American Economic Review*, 97, 586-606.

Tiao G. C., and D. Xu (1993): "Robustness of Maximum Likelihood Estimates for Multi-step Predictions: The Exponential Smoothing Case", *Biometrika*, 80(3), 623-644.

Tonner, J., and J. Bruha (2014): "The Czech Housing Market Through the Lens of a DSGE Model Containing Collateral-Constrained Households", Czech National Bank Working Paper 9/2014.

Xia, Y., and H. Tong (2011): "Feature Matching in Time Series Modeling", *Statistical Science*, 26(1), 21-46.

Appendix A: Data

Table A1. List of variables.

Variable	Description	Units	Seasonal adjustment
RGDP	Real Gross Domestic Product	Index 2000:Q1=100	SAAR
PGDP	Gross Domestic Product: Implicit Price Deflator	Index 2000:Q1=100	SA
Cons	Real Personal Consumption Expenditures	Index 2000:Q1=100	SAAR
GDPInv	Real Gross Private Domestic Investment	Index 2000:Q1=100	SAAR
Emp. Hours	Nonfarm Business Sector: Hours of All Persons	Index 1982:Q1=100	SA
Real Comp/Hour	Nonfarm Business Sector: Real Compensation Per Hour	Index 1982:Q1=100	SA
FedFunds	Effective Federal Funds Rate	%, quarterly average	NSA

Note: SAAR – seasonally adjusted annual rate, SA – seasonally adjusted, NSA – not seasonally adjusted.