Online Appendix: Conditionally Optimal Weights and Forward-Looking Approaches to Combining Forecasts

Christopher G. Gibbs*
and
Andrey L. Vasnev
The University of Sydney, New South Wales, Australia

June 30, 2021

A1 Supplementary material for section 2

A1.1 Proofs

Theorem 1

Proof We prove the optimization formulation. The MSE formulation follows by definition of the optimal solution, and the Jensen-type inequality formulation follows by substituting the explicit solutions.

(a) When comparing the unconditional problem

$$\min_{\boldsymbol{w}} \boldsymbol{w}' \, \mathrm{E}(\boldsymbol{e}_{T+h}\boldsymbol{e}'_{T+h})\boldsymbol{w}$$

and the conditional problem

$$\min_{\boldsymbol{w}} \boldsymbol{w}' \, \mathrm{E}(\boldsymbol{e}_{T+h}\boldsymbol{e}'_{T+h}|I_T)\boldsymbol{w},$$

*Corresponding author contact information: School of Economics, The University of Sydney, christopher.gibbs@sydney.edu.au.

it is useful to consider a function $\psi(\boldsymbol{w}) = \boldsymbol{w}' \operatorname{E}(\boldsymbol{e}_{T+h}\boldsymbol{e}'_{T+h}|I_T)\boldsymbol{w}$. We have

$$\min_{\boldsymbol{w}} \psi(\boldsymbol{w}) \leq \psi(\boldsymbol{v}) \text{ for any } \boldsymbol{v};$$

therefore,

$$\operatorname{E}[\min_{\boldsymbol{w}} \psi(\boldsymbol{w})] \leq \operatorname{E}[\psi(\boldsymbol{v})]$$

and

$$\operatorname{E}[\min_{\boldsymbol{w}} \psi(\boldsymbol{w})] \leq \min_{\boldsymbol{v}} \operatorname{E}[\psi(\boldsymbol{v})],$$

which is equivalent to

$$\operatorname{E}[\min_{\boldsymbol{w}} \psi(\boldsymbol{w})] \leq \min_{\boldsymbol{w}} \operatorname{E}[\psi(\boldsymbol{w})].$$

In the original notation, we have

$$\operatorname{E}\left[\min_{\boldsymbol{w}} \boldsymbol{w}' \operatorname{E}(\boldsymbol{e}_{T+h}\boldsymbol{e}'_{T+h}|I_T)\boldsymbol{w}\right] \leq \min_{\boldsymbol{w}} \operatorname{E}\left[\boldsymbol{w}' \operatorname{E}(\boldsymbol{e}_{T+h}\boldsymbol{e}'_{T+h}|I_T)\boldsymbol{w}\right],$$

which provides us with

$$\operatorname{E}\left[\min_{\boldsymbol{w}} \boldsymbol{w}' \operatorname{E}(\boldsymbol{e}_{T+h}\boldsymbol{e}'_{T+h}|I_T)\boldsymbol{w}\right] \leq \min_{\boldsymbol{w}} \boldsymbol{w}' \operatorname{E}(\boldsymbol{e}_{T+h}\boldsymbol{e}'_{T+h})\boldsymbol{w}.$$

The optimization problems have explicit solutions; thus,

$$\operatorname{E}\left[\frac{1}{\boldsymbol{\iota}' \left[\operatorname{E}(\boldsymbol{e}_{T+h}\boldsymbol{e}'_{T+h}|I_T)\right]^{-1}\boldsymbol{\iota}}\right] \leq \frac{1}{\boldsymbol{\iota}' \left[\operatorname{E}(\boldsymbol{e}_{T+h}\boldsymbol{e}'_{T+h})\right]^{-1}\boldsymbol{\iota}}$$

or

$$\operatorname{E}\left(\frac{1}{\boldsymbol{\iota}'[\Sigma_\xi + \boldsymbol{b}_T\boldsymbol{b}'_T]^{-1}\boldsymbol{\iota}}\right) \leq \frac{1}{\boldsymbol{\iota}'\Sigma_e^{-1}\boldsymbol{\iota}}.$$

In other words, using the predictability of the forecast errors is beneficial because we achieve a lower MSE in expectation.

(b) The proof is similar to part (a) and simply requires substituting the unconditional expectation with the expectation conditional on J_T.

(c) The proof follows from part (b) if $\psi(\boldsymbol{w}) = \mathrm{E}[L(e_{c,T+h})|I_T]$. ∎

Theorem 2

Proof There are several critical elements that are affected by increasing γ. Two simple terms are $\Sigma_\eta = \gamma^2\Sigma_{\eta_0}$ and $\widehat{\boldsymbol{b}}_T = \boldsymbol{b} + \gamma\boldsymbol{\eta}_0$. For other terms, $\mathrm{E}(\boldsymbol{\eta}_T|\widehat{\boldsymbol{b}}_T)$, $\mathrm{var}(\boldsymbol{\eta}_T|\widehat{\boldsymbol{b}}_T)$ in $\mathrm{MSE}(\boldsymbol{w}^{\dagger\dagger})$ and $\mathrm{E}(\boldsymbol{b}_T|\widehat{\boldsymbol{b}}_T)$, $\mathrm{var}(\boldsymbol{b}_T|\widehat{\boldsymbol{b}}_T)$ in $\mathrm{MSE}(\boldsymbol{w}(\widehat{\boldsymbol{b}}_T))$, we need to know the conditional moments.

Without loss of generality, we prove the theorem in a special when $\boldsymbol{b}_T$ and $\boldsymbol{\eta}_T$ are normal and independent, i.e.

$$\begin{pmatrix} \boldsymbol{b}_T \\ \boldsymbol{\eta}_T \\ \widehat{\boldsymbol{b}}_T \end{pmatrix} \sim N\left(\boldsymbol{0}, \begin{pmatrix} \Sigma_b & \boldsymbol{0} & \Sigma_b \\ \boldsymbol{0} & \Sigma_\eta & \Sigma_\eta \\ \Sigma_b & \Sigma_\eta & \Sigma_b + \Sigma_\eta \end{pmatrix}\right).$$

The conditional distributions are

$$\boldsymbol{b}_T|\widehat{\boldsymbol{b}}_T \sim N(\Sigma_b(\Sigma_b + \Sigma_\eta)^{-1}\widehat{\boldsymbol{b}}_T, \Sigma_b - \Sigma_b(\Sigma_b + \Sigma_\eta)^{-1}\Sigma_b)$$

and

$$\boldsymbol{\eta}_T|\widehat{\boldsymbol{b}}_T \sim N(\Sigma_\eta(\Sigma_b + \Sigma_\eta)^{-1}\widehat{\boldsymbol{b}}_T, \Sigma_\eta - \Sigma_\eta(\Sigma_b + \Sigma_\eta)^{-1}\Sigma_\eta).$$

They are affected by increasing γ in the following way.

$$\mathrm{E}(\boldsymbol{b}_T|\widehat{\boldsymbol{b}}_T) = \Sigma_b(\Sigma_b + \gamma^2\Sigma_{\eta_0})^{-1}(\boldsymbol{b} + \gamma\boldsymbol{\eta}_0) \to \boldsymbol{0},$$

$$\mathrm{var}(\boldsymbol{b}_T|\widehat{\boldsymbol{b}}_T) = \Sigma_b - \Sigma_b(\Sigma_b + \gamma^2\Sigma_{\eta_0})^{-1}\Sigma_b \to \Sigma_b,$$

$$\mathrm{E}(\boldsymbol{\eta}_T|\widehat{\boldsymbol{b}}_T) = \gamma^2\Sigma_{\eta_0}(\Sigma_b + \gamma^2\Sigma_{\eta_0})^{-1}(\boldsymbol{b} + \gamma\boldsymbol{\eta}_0) \to \infty,$$

$$\mathrm{var}(\boldsymbol{\eta}_T|\widehat{\boldsymbol{b}}_T) = \gamma^2\Sigma_{\eta_0} - \gamma^2\Sigma_{\eta_0}(\Sigma_b + \gamma^2\Sigma_{\eta_0})^{-1}\gamma^2\Sigma_{\eta_0} \to \boldsymbol{0}.$$

We can now see that

$$\mathrm{MSE}(\boldsymbol{w}^{\dagger\dagger}) = \frac{\boldsymbol{\iota}'[\Sigma_\xi + \gamma^2\Sigma_{\eta_0}]^{-1}\left[\Sigma_\xi + \mathrm{var}(\boldsymbol{\eta}_T|\widehat{\boldsymbol{b}}_T) + \mathrm{E}(\boldsymbol{\eta}_T|\widehat{\boldsymbol{b}}_T)\,\mathrm{E}(\boldsymbol{\eta}_T'|\widehat{\boldsymbol{b}}_T)\right][\Sigma_\xi + \gamma^2\Sigma_{\eta_0}]^{-1}\boldsymbol{\iota}}{(\boldsymbol{\iota}'[\Sigma_\xi + \gamma^2\Sigma_{\eta_0}]^{-1}\boldsymbol{\iota})^2} \to \infty$$

and $\mathrm{MSE}(\boldsymbol{w}(\widehat{\boldsymbol{b}}_T)) =$

$$= \frac{\boldsymbol{\iota}'[\Sigma_\xi + \gamma^2\Sigma_{\eta_0} + (\boldsymbol{b}+\gamma\boldsymbol{\eta}_0)(\boldsymbol{b}+\gamma\boldsymbol{\eta}_0)']^{-1}\left[\Sigma_\xi + \mathrm{var}(\boldsymbol{b}_T|\widehat{\boldsymbol{b}}_T) + \mathrm{E}(\boldsymbol{b}_T|\widehat{\boldsymbol{b}}_T)\,\mathrm{E}(\boldsymbol{b}_T'|\widehat{\boldsymbol{b}}_T)\right][\dots]^{-1}\boldsymbol{\iota}}{(\boldsymbol{\iota}'[\Sigma_\xi + \gamma^2\Sigma_{\eta_0} + (\boldsymbol{b}+\gamma\boldsymbol{\eta}_0)(\boldsymbol{b}+\gamma\boldsymbol{\eta}_0)']^{-1}\boldsymbol{\iota})^2}$$

$$\to \frac{\boldsymbol{\iota}'[\Sigma_{\eta_0} + \boldsymbol{\eta}_0\boldsymbol{\eta}_0']^{-1}\,[\Sigma_\xi + \Sigma_b]\,[\Sigma_{\eta_0} + \boldsymbol{\eta}_0\boldsymbol{\eta}_0']^{-1}\boldsymbol{\iota}}{(\boldsymbol{\iota}'[\Sigma_{\eta_0} + \boldsymbol{\eta}_0\boldsymbol{\eta}_0']^{-1}\boldsymbol{\iota})^2}.$$

This proves the theorem for normal and independent $\boldsymbol{b}_T$ and $\boldsymbol{\eta}_T$. If $\boldsymbol{b}_T$ and $\boldsymbol{\eta}_T$ are not independent, i.e., $\mathrm{cov}(\boldsymbol{b}_T, \boldsymbol{\eta}_T) = \Sigma_{b\eta}$, then it will be affected by increasing γ, but the effect is linear, i.e., $\Sigma_{b\eta} = \gamma\Sigma_{b\eta_0}$, and it will be dominated by quadratic terms, such as $\gamma^2\Sigma_{\eta_0}$ and $\gamma^2\boldsymbol{\eta}_0\boldsymbol{\eta}_0'$.

Finally, we observe that in the case of the elliptical distribution for $(\boldsymbol{b}_T, \boldsymbol{\eta}_T, \widehat{\boldsymbol{b}}_T)'$, the formulae for the conditional mean and the conditional variance remain the same, so the result holds in the general case. ∎

Theorem 3

Proof The proof follows from the fact that the relationship between

$$\mathrm{MSE}(\boldsymbol{w}^{\dagger\dagger}) = \frac{1}{\boldsymbol{\iota}'[\Sigma_\xi + \Sigma_\eta]^{-1}\boldsymbol{\iota}}$$

and

$$\mathrm{MSE}(\boldsymbol{w}^*) = \frac{1}{\iota'[\Sigma_\xi + \mathrm{E}(\boldsymbol{b}_T\boldsymbol{b}_T')]^{-1}\iota}$$

is the same as the relationship between Σ_η and $\mathrm{E}(\boldsymbol{b}_T\boldsymbol{b}_T')$. ∎

A1.2 Specific models for b_T

A1.2.1 MA(1) model

If $b_T = \theta\xi_T$, then Theorem 1 guaranties that $\mathrm{E}(\mathrm{MSE}(\boldsymbol{w}^*(I_T))) \leq \mathrm{MSE}(w^*)$, which is equivalent to

$$\mathrm{E}\left(\frac{1}{\boldsymbol{\iota}'[\Sigma_\xi + \theta^2\xi_T\xi_T']^{-1}\boldsymbol{\iota}}\right) \leq \frac{1+\theta^2}{\boldsymbol{\iota}'\Sigma_\xi^{-1}\boldsymbol{\iota}}.$$

A stronger inequality without the expectation can hold in certain cases. For example, due to positive definite $\theta^2 \xi_T \xi_T'$,

$$\frac{1}{\boldsymbol{\iota}'\Sigma_\xi^{-1}\boldsymbol{\iota}} < \frac{1}{\boldsymbol{\iota}'[\Sigma_\xi + \theta^2 \xi_T \xi_T']^{-1}\boldsymbol{\iota}}.$$

However if the left hand side is multiplied by $(1+\theta^2)$, it can be

$$\frac{1}{\boldsymbol{\iota}'[\Sigma_\xi + \theta^2 \xi_T \xi_T']^{-1}\boldsymbol{\iota}} < \frac{1+\theta^2}{\boldsymbol{\iota}'\Sigma_\xi^{-1}\boldsymbol{\iota}}$$

depending on the parameters.

A1.2.2 AR(1) model

Similarly, if $b_T = \phi \mathrm{e}_T$, then Theorem 1 guaranties that $\mathrm{E}(\mathrm{MSE}(\boldsymbol{w}^*(I_T))) \leq \mathrm{MSE}(w^*)$, which is equivalent to

$$\mathrm{E}\left(\frac{1}{\boldsymbol{\iota}'[\Sigma_\xi + \phi^2 \mathrm{e}_T \mathrm{e}_T']^{-1}\boldsymbol{\iota}}\right) \leq \frac{1}{1-\phi^2}\frac{1}{\boldsymbol{\iota}'\Sigma_\xi^{-1}\boldsymbol{\iota}}.$$

Again, the positive definiteness of $\phi^2 \mathrm{e}_T \mathrm{e}_T'$,

$$\frac{1}{\boldsymbol{\iota}'\Sigma_\xi^{-1}\boldsymbol{\iota}} < \frac{1}{\boldsymbol{\iota}'[\Sigma_\xi + \phi^2 \mathrm{e}_T \mathrm{e}_T']^{-1}\boldsymbol{\iota}},$$

but the division with $(1-\phi^2)$ can produce

$$\frac{1}{\boldsymbol{\iota}'[\Sigma_\xi + \phi^2 \mathrm{e}_T \mathrm{e}_T']^{-1}\boldsymbol{\iota}} < \frac{1}{1-\phi^2}\frac{1}{\boldsymbol{\iota}'\Sigma_\xi^{-1}\boldsymbol{\iota}}$$

depending on the parameters.

One example is when $b_T = (\phi e_{T,1}, 0 \ldots 0)'$, i.e., only the first element follows an AR(1) process while the other forecasts are conditionally unbiased, and $\Sigma_\xi = \sigma_\xi^2 I$, i.e., the forecasts have the same variance and uncorrelated with each other. In this case a stronger version of Theorem 1, i.e., the inequality without expectation, will hold if $\phi(1-\phi) < \sigma_\xi^2$.

A1.3 Monte Carlo study

We put forward three Monte Carlo exercises to numerically illustrate the point that noisy correction leads to decreasing forecast accuracy while using that same information to correct weights does not. The first two are simple Monte Carlo exercises in an i.i.d. environment. The first considers an omitted variable bias where we vary the importance of that omitted variable to the data generating process. The second demonstrates how bias correction and conditionally optimal weights change as we vary the signal to noise ratio in the conditional bias when only a noisy signal is available.

For the third exercise, we consider a more complicated data generating process of a conditional location-scale model and follow Lima and Meng (2017) to investigate the forecasting power of conditionally optimal weights in a weak predictor environment, where there are many competing forecasting models. This environment also allows us to explore the properties of conditionally optimal weights when we vary the number of combined forecasts, n.

A1.3.1 Omitted variable bias

The data generating process (DGP) we consider follows

$$y_{t+1} = x_t + z_t + \mu_{t+1}, \tag{A1}$$

where x_t and z_t are independent and follow i.i.d. mean zero normal random processes with unit variance and μ_{t+1} is error term whose variance we will vary. We assume that the forecaster does not know the DGP and considers the three following model specifications to form one-step-ahead forecasts: regression of y_{t+1} on x_t that produces forecast $f_{1,t+1}$; regression of y_{t+1} on z_t that produces forecast $f_{2,t+1}$; regression of y_{t+1} on x_t and z_t that produces forecast $f_{3,t+1}$. This setup provides an environment where the unconditional optimal weights out-perform the equally weighted forecast.

We assume there is an omitted variable such that μ_{t+1} is composed of two components

$\mu_{t+1} = b_t + \zeta_{t+1}$, where b_t is i.i.d. and mean zero but available to the forecasters and ζ_t has unit variance. While forecasters are sure that x_t and z_t are useful for forecasting y_t, they are uncertain about b_t. We reinforce this by assuming that the variance of b_t is small relative to x_t and z_t. We then compare whether b_t is more useful to include to bias-correct or to construct conditionally optimal weights. Specifically, we compare (1) equal weights, (2) the unconditional optimal weights as in Bates and Granger (OW), (3) the conditionally optimal weights constructed using b_t (COW), and (4) the classical optimal weights for the bias-corrected forecasts constructed using b_t (BC-OW). In the last strategy, we first attempt to remove the bias of the individual forecasts using b_t and then construct unconditional optimal combination weights based on the biased-corrected forecasts as described in Section 2.2.

We compare the combined forecast by conducting a standard pseudo-out-of-sample forecasting exercise, where we partition the simulated data into in-sample and out-of-sample subsets and recursively forecast the out-of-sample subset by re-estimating the forecast models and recalculating the weights in each period. To mimic how one would implement estimated weights in practice, the in-sample period is partitioned into two separate subperiods: 20 periods to obtain initial estimates by OLS of the forecast model parameters; and 20 periods to obtain initial recursive forecast errors for the three models to construct initial estimates of the bias and to construct initial estimates of the weights. We use the following model for bias prediction:

$$y_{t+1} - f_{i,t+1} = c_i + \beta_i b_t + \epsilon_{i,t+1},$$

where $i = 1, 2$, or 3. We run 5,000 simulations for each case.

Figure A1 shows the results for recursive out-of-sample forecasts for 25 and 50 periods (T). From the graph we can see that when the variance of the bias is small relative to the DGP that bias-corrected forecasts are the least accurate forecasts (recall x_t, z_t, and ζ_t each have unit variance). In the small samples we consider, when the bias is small, there

is not enough power to reject the null hypothesis $\beta_i = 0$ at conventional significance levels. Attempts to correct the bias using b_t in these cases introduces more noise to the forecasts and accuracy suffers. However, as one would expect, as the importance of b_t grows, it becomes more useful for forecasting. If the variance is large enough, the forecast suffers greatly by not including b_t in the forecast. Conditionally optimal weights, however, provide a uniform forecast throughout the sample. Because any noise in the bias predictions balance in the numerator and the denominator of the weights, the forecasts are less affected when bias is small, and predictions of the bias are inaccurate.

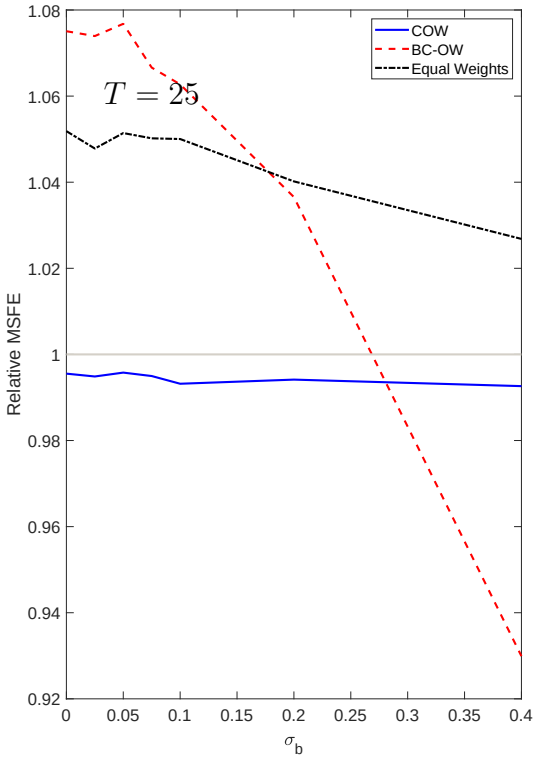
Lastly, the conditionally optimal weights are applied to the uncorrected forecasts, which omit b_t. Therefore, there is a clear limit to how accurate the combined forecasts can be because b_t is a driver of the DGP. The advantage of conditionally optimal weights is that one can use b_t even when it has a tenuous relationship to the DGP to improve forecast accuracy in finite samples without risking a significant loss in forecast accuracy.

A1.3.2 Signal to noise ratio of the bias

In this scenario, we add bias to this forecasting problem in the same way described in Section 2.2 by assuming that μ_{t+1} is composed of two components: $\mu_{t+1} = b_t + \zeta_{t+1}$, of which only a noisy signal of b_t, $\hat{b}_t = b_t + \eta_t$, is available. Therefore, whether $\hat{b}_t$ is useful for forecasting y_{t+1} will depend on the nature η_t. One interpretation of η_t is that it represents measurement issues such as with GDP and its components, which are themselves based on estimates that undergo frequent revisions.

We assume that b_t, ζ_t and η_t are independent from each other and follow i.i.d normal random processes. We then fix ζ_t to have a unit variance and vary the relative variances of b_t and η_t and the size of the out-of-sample period (T) to compare out-of-sample mean square forecast errors for four different combination strategies: (1) equal weights, (2) the unconditional optimal weights as in Bates and Granger (OW), (3) the conditionally optimal weights constructed using $\hat{b}_t$ (COW), and (4) the classical optimal weights for the bias-corrected forecasts constructed using $\hat{b}_t$ (BC-OW).

Figure A1: Monte Carlo Simulation



T=25			
σ_b	COW	EW	BC-OW
0.01	0.996	1.052	1.075
0.03	0.995	1.048	1.074
0.05	0.996	1.051	1.077
0.08	0.995	1.050	1.067
0.10	0.993	1.050	1.063
0.20	0.994	1.040	1.037
0.40	0.993	1.027	0.930

T=50			
σ_b	COW	EW	BC-OW
0.01	0.998	1.098	1.058
0.03	0.998	1.092	1.058
0.05	0.998	1.096	1.058
0.08	0.998	1.094	1.052
0.10	0.997	1.095	1.047
0.20	0.997	1.087	1.018
0.40	0.999	1.070	0.915

Notes: Monte Carlo simulation comparing relative MSFE of bias-corrected forecasts combined with optimal weights (BC-OW), conditionally optimal weights forecasts (COW), and equal weights (EW). MSFE are shown relative to an optimal weight (OW) forecast.

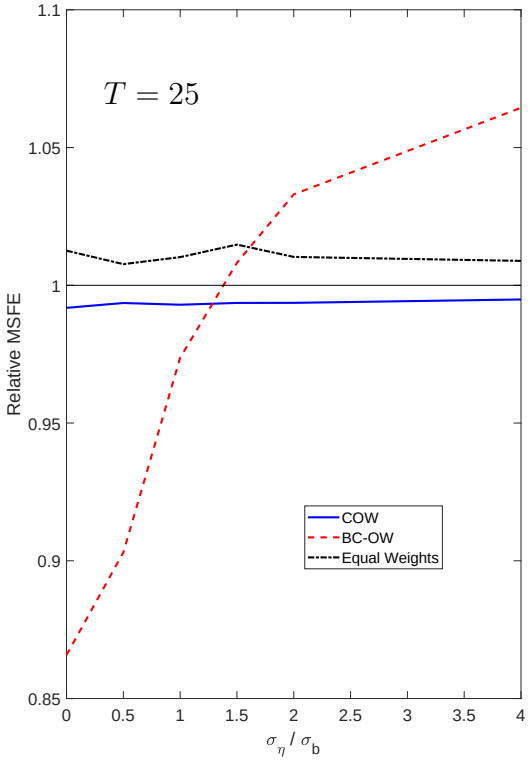
We again compare the combined forecast by conducting a standard pseudo-out-of-sample forecasting exercise, where we partition the simulated data into in-sample and out-of-sample subsets and recursively forecast the out-of-sample subset by re-estimating the forecast models and recalculating the weights in each period. We use the same model for bias prediction as in previous section A1.3.1 and we again run 5,000 simulations for each case.

Figure A2 summarizes the results. When there is no noise, BC-OW yields the lowest mean squared forecast error (MSFE). But as the variance of the noise increases, the forecast accuracy of this strategy falls, and COW becomes the best performing strategy. Because we assume that x_t, z_t, b_t, and ζ_t are uncorrelated, the bias affects each model the same way in the limit, which as we have shown in Section 2 of the paper, means that conditionally optimal weights and unconditional optimal weights are asymptotically the same. However,

we find that in small samples with recursive estimation, COW delivers a small improvement over OW. As the out-of-sample period T becomes large, though, the MSFE of COW and OW forecasts converge in this special case.

This exercise, though, is not very informative about the true forecasting power of this approach because optimal weights easily outperform equal weights, which clearly does not reflect reality when these strategies are used in practice. Therefore, we rely on real-time out-of-sample forecasting evaluations to illustrate the main advantages of the proposed strategy.

Figure A2: Monte Carlo Simulation



$T = 25$			
σ_η/σ_b	COW	EW	BC-OW
0	0.992	1.013	0.866
0.5	0.994	1.008	0.903
0.75	0.993	1.010	0.974
1	0.994	1.015	1.008
2	0.994	1.010	1.033
4	0.995	1.009	1.064
8	0.996	1.015	1.075

$T = 50$			
σ_η/σ_b	COW	EW	BC-OW
0	0.999	1.057	0.850
0.5	0.999	1.052	0.890
0.75	0.997	1.055	0.957
1	0.998	1.056	0.994
2	0.997	1.056	1.015
4	0.997	1.054	1.047
8	0.997	1.057	1.056

Notes: Monte Carlo simulation comparing relative MSFE of bias-corrected forecasts combined with optimal weights (BC-OW), conditionally optimal weights forecasts (COW), and equal weights (EW). MSFE are shown relative to an optimal weight (OW) forecast.

A1.4 Forecasting in a more complex environment

We now turn to a more complex data generating process to explore conditionally optimal weights and some of the proposed alternative formulations from Section 2.3 in an environment

with weak and partially weak predictors. For this exercise, we follow Lima and Meng (2017) and consider the following location-scale model:

$$\begin{aligned} y_{t+1} &= \beta_0 + \sum_i \beta_i x_{i,t} + \left(\gamma_0 + \sum_i \gamma_i x_{i,t}\right) \eta_{t+1} \\ & i = 1, 2, 3, ..., 6; \; t = 1, 2, ..., 1000, \end{aligned} \tag{A2}$$

where we assume $\beta_0 = 1$, $\eta_{t+1} \sim N(0, \sigma_\eta^2)$, and $\sigma_\eta = 0.75$. The sample size is set to 1000. For pseudo-forecasting purposes, we partition the sample into three subsets: 1) $t < 500$, 2) $500 \leq t < 901$, and 3) $t > 900$. The first subset is the in-sample period that is used to provide initial estimates for our individual forecast models, which we describe later. The second subset is used for in-sample recursive forecasting to generate a sample of forecast errors for each individual forecast model. The errors are used to create the conditional bias estimates required for conditionally optimal weights. The final subset is the pseudo out-of-sample period, which we recursively forecast with the individual models and forecast combinations to evaluate the efficacy of the forecasting strategies of interest.

The number of potential predictor, $x_{i,t}$, is fixed at six. We assume the predictors are drawn from a uniform distributions over $(0, 1)$, where, as in Lima and Meng (2017), we consider Spearmann correlation among the predictors of $\rho_{i,j} = (0, 0.1, 0.25, 0.5, 0.95)$. We assume that the forecasters have access to all six variables and that they consider parsimonious linear models that include different combinations of the six predictors. Specifically. following Elliott et al. (2013) and our Section 3.3, we consider all distinct subsets of the six predictors of size 1, 2, and 3. This leads to 41 models of the form

$$y_{t+1} = b_0 + b_i x_{i,t} + b_j x_{j,t} + b_k x_{k,t} + e_t$$

where for subsets of size one, $b_j = b_k = 0$ and $i = 1, ..., 6$; for subsets of size two, $b_k = 0$ and there are 15 different combinations of $x_{i,t}$ and $x_{j,t}$; and for subsets of size 3 we have 20 different combinations of $x_{i,t}, x_{j,t}$ and $x_{k,t}$.

We then study combined forecasts of the 6 individual models ($k = 1$ in the notation of Complete Subset Regression (CSR) of Elliott et al., 2013), 21 models comprised of all subsets of $n \leq 2$ ($k = 2$), and 41 models comprised of all subsets of $n \leq 3$ ($k = 3$). The combination strategies we consider are:

1. Conditionally Optimal Weights (see Section 2.1 of the main text)
2. Bias-corrected Optimal Weights (see Section 2.2 of the main text)
3. Bias-corrected Equal Weights (see Section 2.2 of the main text)
4. Predicted Exponential Weights with $\gamma = 1$ and 10 (see Section 2.3 of the main text)
5. Equal Weights (CSR).

For the COW and PEW forecasts, we model conditional bias as an AR(1) with a constant:

$$\hat{y}_{i,t+1} - y_{t+1} = e_{i,t+1} = c_i + \rho e_{i,t} + \zeta_{t+1}.$$

The univarate benchmark is a simple recursive average of y_t.

The Monte Carlo experiment assumes weak and partially weak predictors as in the baseline case considered by Lima and Meng with $\beta_i = \gamma_i = 0$ for $i = 3, 4, ..., 6$ for all t, $\beta_1 = -1.5$ and $\gamma = 5$ if $\eta_{t+1} \leq \phi^{-1}(0.5)$, and $\beta_2 = 1.5$ and $\gamma_2 = 5$ if $\eta_{t+1} > \phi^{-1}(0.5)$, where $\phi^{-1}(x)$ refers to $x \times 100$ percentile of the distribution of η_t. Finally, to add outliers it is assumed that $\gamma_0 = \sigma_\eta$ if $\eta_{t+1} < 1.96$ and $\gamma_0 = 5\sigma_\eta$ if $\eta_{t+1} > 1.96$.

Table A1 shows the mean Monte Carlo results from 250 experiments that each include 100 out-of-sample forecasts. Therefore, each value shown summarizes 25,000 out-of-sample forecasts. All results are reported relative to the univariate benchmark with the average Clark and West (2007) test statistic from the 250 experiments shown on the right. Note that by construction only $x_{1,t}$ and $x_{2,t}$ have any ability to forecast y_{t+1}. The remaining predictors are noise and provide no forecasting power above the benchmark. The forecasting power of these two predictors wanes as the correlation among the predictors rises. The reason for this

decline is over-fitting of the data as it becomes more difficult to distinguish which movements in $x_{1,t}$ and $x_{2,t}$ are responsible for changes in the DGP.

The combined forecasts all show some ability to forecast y_{t+1} relative to the the benchmark forecast. However, the same pattern with respect to the correlation of the predictors remains. When the predictors provide distinct information, the combined forecasts do well. When the predictors become similar, over-fitting occurs and the combined forecasts lose forecast accuracy.

Looking across the different forecast combination strategies, there is clear relationship with combined forecast accuracy and whether the combination strategy nests optimal weights or equal weights. The former relies on a variance-covariance estimate of the forecast errors, while the latter specifications do not use this information. The first pattern is that optimal methods perform best when the number of models combined (n) is small and the correlation among the predictors is low. For example, COW and BC-OW combinations of the six single $x_{i,t}$ prediction models provides the lowest MSFE among the strategies tested when $\rho \leq 0.25$. In contrast, when $\rho \geq 0.5$ the best strategies rely on equal weights and combine all 41 models.

The explanation for this pattern is straightforward. The majority of the forecasts that are combined in this exercise are noise. Only two of the six predictors contain any information about the DGP by construction. The rest are pure noise. Averaging over that noise with CSR, or BC-EW, zeroes it out and leads to better forecast accuracy. With COW or BC-OW, however, the variance-covariance of the errors from the forecasts is exploited. The addition of many nearly identical noise forecasts leads to highly correlated forecast errors. The highly correlated forecast errors lead to near multicolinearity in the forecast error variance-covariance estimate and instability in the estimated weights when estimated recursively over time.

Conditionally optimal weights, however, does not rely solely on the variance-covariance of past forecast errors. It also makes use of the estimated conditional bias. Predicted exponential weights (PEW) makes use of just the conditional bias prediction to inform the weights. Therefore, the method loses the information obtained in the variance-covariance

of past forecast errors, but it removes the multicolinearity issue. Here weights are adjusted to put the most weight on the forecast with the smallest predicted squared bias in each period. The parameter γ provides shrinkage to the weights by redistributing more or less weight to the best models depending on its predicted bias. When γ is small, the weights are close to equal weights. When γ is large, nearly all weight is placed on a single model. This explains why the $\gamma = 1$ case is nearly identical to the CSR results. The weights are constrained to be very close to the equal weights. In the $\gamma = 10$ case, the weights have much more variation and more weight is placed on the forecasts with the lowest expected bias. It is the conditional bias estimates that drives the consistent improvement in forecast accuracy observed here relative to the CSR results. The gains are modest because there is not much conditional bias to exploit among the considered models and the chosen DGP.

The last thing to note from this exercise is that the combinations which bias-correct the forecasts before combining forecasts systematically yield higher MSFE than the combinations that use the predicted bias to construct weights of the uncorrected forecasts. This illustrates the insights discussed in Section 2.2. Bias-correcting the individual forecasts can introduce noise that the combined forecast cannot remove. However, that same information can be used to construct combination weights that improve upon other combination methods such as CSR.

A2 Supplementary material for Section 3

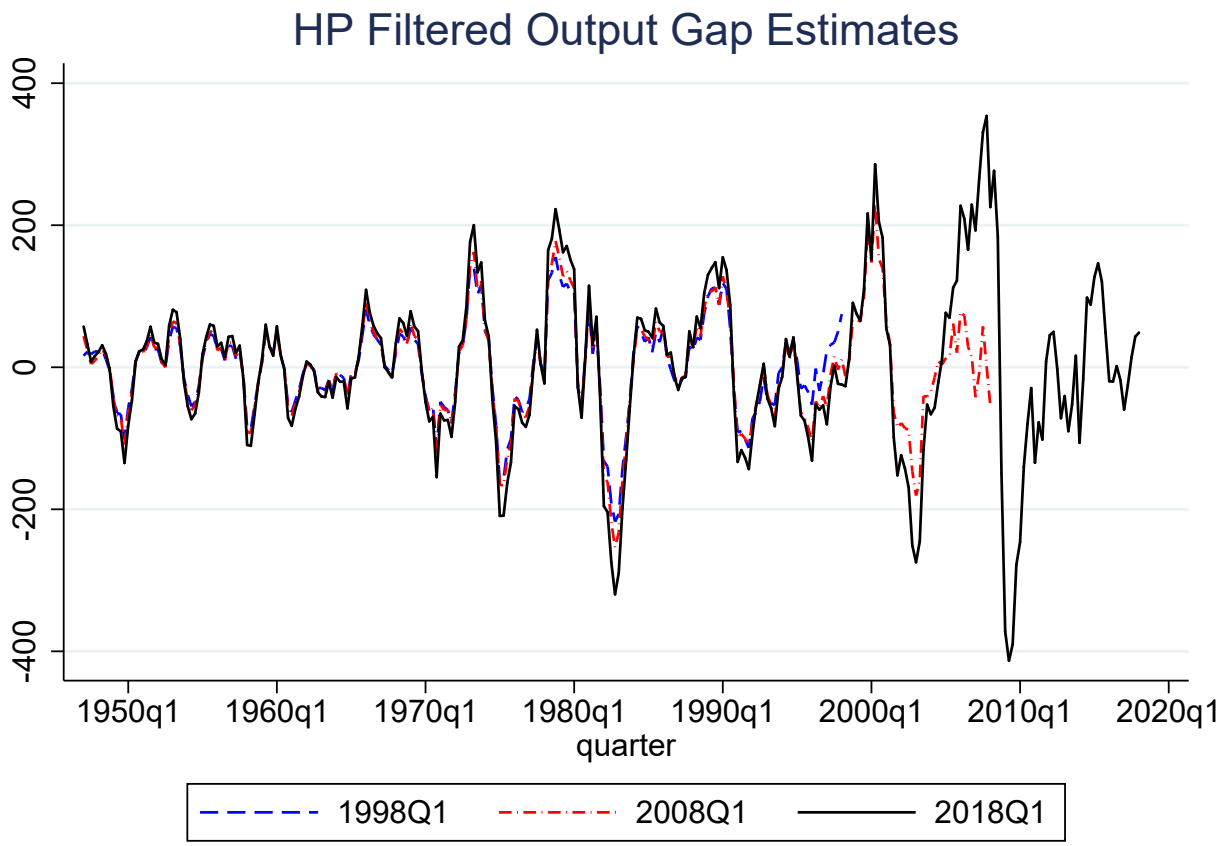
We use real time data for the forecasting exercise. This means that we estimate all of our models anew each quarter using the most up-to-date data available in that quarter. Therefore, over time, the information that informs the forecasts can vary greatly. The most affected measure we consider is the HP filtered output gap. Figure A3 shows the significant variation that this procedure introduces. The HP filter is one-sided at the end of the sample which leads to substantial revisions once new data is available. The process of re-estimating the models anew each quarter ensures that we are never evaluating a forecast that a forecaster

Table A1: Monte Carlo Results for Scale-Location DGP - Weak and Partially Weak Predictors

	$\rho = 0$		$\rho = 0.1$		$\rho = 0.25$		$\rho = 0.5$		$\rho = 0.95$	
Model	Rel. MSFE	CW-stat	Rel. MSFE	CW-stat	Rel. MSFE	CW-stat	Rel. MSFE	CW-stat	Rel. MSFE	CW-stat
x_1	0.9637	1.92	0.9730	1.65	0.9833	1.31	0.9951	0.74	1.0015	-0.27
x_2	0.9641	1.92	0.9733	1.66	0.9815	1.37	0.9923	0.84	1.0016	-0.25
x_3	1.0012	-0.12	1.0012	-0.14	1.0006	-0.22	1.0009	-0.15	1.0015	-0.23
x_4	1.0006	-0.21	1.0009	-0.23	1.0011	-0.29	1.0011	-0.22	1.0014	-0.25
x_5	1.0015	-0.28	1.0015	-0.31	1.0012	-0.18	1.0018	-0.22	1.0016	-0.24
x_6	1.0011	-0.24	1.0011	-0.31	1.0011	-0.15	1.0016	-0.21	1.0015	-0.27
EW - 6/ CSR (k=1)	0.9757	2.73	0.9816	1.64	0.9881	1.38	0.9957	0.98	1.0014	-0.26
CSR (k=2)	0.9615	2.73	0.9699	2.46	0.9786	2.09	0.9883	1.43	1.0008	0.03
CSR (k=3)	0.9511	2.73	0.9610	2.47	0.9712	2.13	0.9830	1.56	1.0008	0.09
COW- 6	0.9366	2.65	0.9491	2.40	0.9641	2.04	0.9851	1.39	1.0138	-0.12
COW - 21	0.9694	2.27	0.9868	2.03	1.0062	1.57	1.0260	1.14	1.0599	0.12
COW - 41	1.0024	2.05	1.0290	1.71	1.0526	1.32	1.0662	0.88	1.1153	0.00
PEW - 6 ($\gamma = 1$)	0.9757	2.73	0.9816	2.45	0.9881	1.99	0.9957	1.03	1.0014	-0.26
PEW - 21 ($\gamma = 1$)	0.9615	2.73	0.9699	2.46	0.9785	2.09	0.9883	1.43	1.0008	0.03
PEW - 41 ($\gamma = 1$)	0.9512	2.73	0.9610	2.47	0.9711	2.13	0.9830	1.56	1.0008	0.09
PEW - 6 ($\gamma = 10$)	0.9754	2.59	0.9814	2.32	0.9875	1.93	0.9955	1.01	1.0014	-0.25
PEW - 21 ($\gamma = 10$)	0.9615	2.65	0.9698	2.39	0.9780	2.06	0.9881	1.42	1.0009	0.03
PEW - 41 ($\gamma = 10$)	0.9514	2.68	0.9612	2.42	0.9708	2.12	0.9829	1.55	1.0008	0.09
BC-OW - 6	0.9393	2.62	0.9608	2.18	0.9623	2.02	0.9875	1.40	1.0195	-0.04
BC-OW - 21	0.9706	2.28	1.0048	1.74	1.0101	1.59	1.0281	1.13	1.0720	0.08
BC-OW - 41	1.0166	1.95	1.0596	1.45	1.0645	1.26	1.0891	0.83	1.1308	0.04
BC-EW - 6	0.9797	1.81	0.9885	1.33	0.9917	1.05	0.9986	0.45	1.0056	-0.09
BC-EW - 21	0.9651	2.28	0.9769	1.79	0.9817	1.57	0.9907	1.02	1.0043	0.05
BC-EW - 41	0.9544	2.47	0.9681	2.01	0.9738	1.82	0.9851	1.29	1.0036	0.14
# Simulations:	250		250		250		250		250	

Notes: Monte Carlo simulations for combined forecasts of a location-scale model following Lima and Meng (2017) given by equation A2. Each simulation recursively forecasts 100 periods and the MSFE is recorded. The Clark and West (2007) test statistic (CW-stat) for equal forecast accuracy of the combined forecast relative to a simple average of past observations is calculated for the 100 forecasts. The table shows the mean relative MSFE and mean CW-stat for 250 simulations.

Figure A3: Real-time data used for forecasting



Notes: HP filtered estimates of the output gap using different vintages of real GDP.

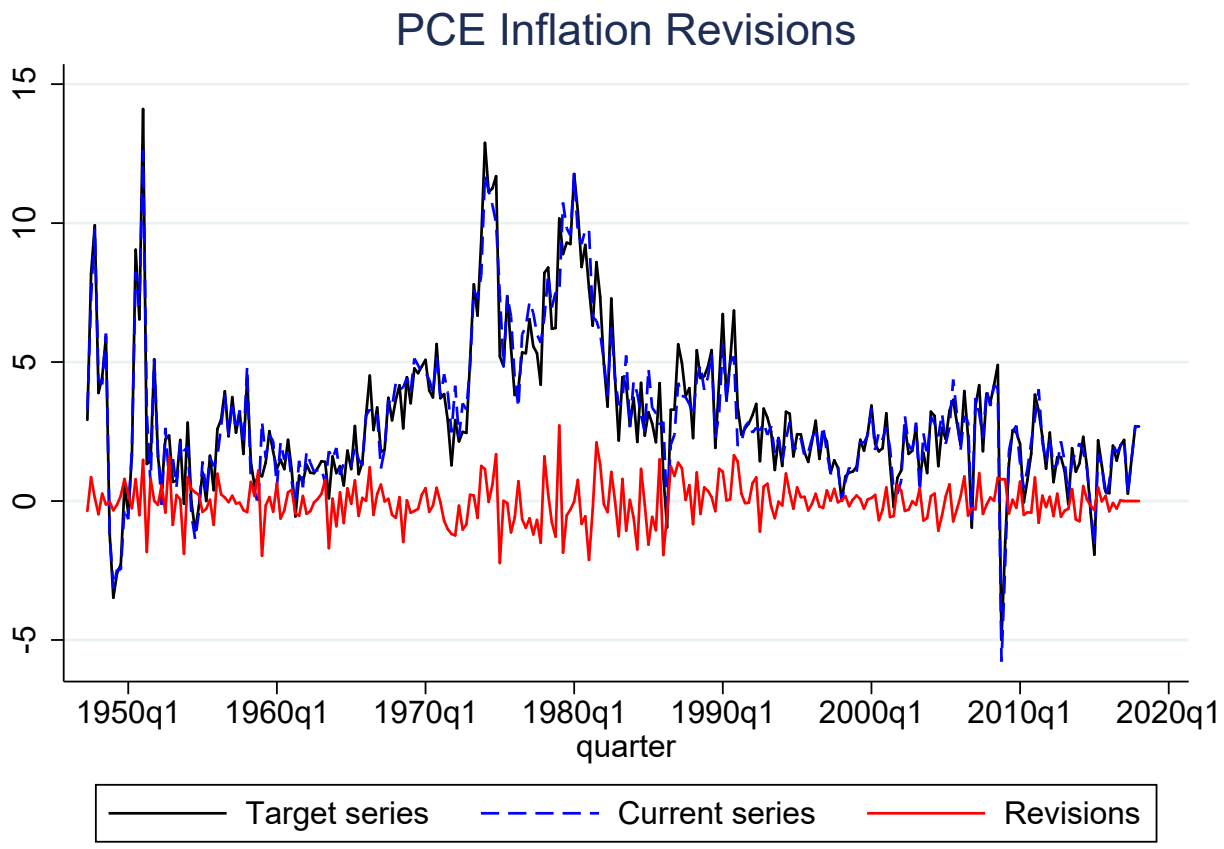
could not have constructed at that time.

Figure A4 shows how the revisions affect our target series: PCE inflation. We follow the standard practice in the literature of using a composite series of the second releases of PCE inflation to compare to the real-time forecasts. This is because some revisions to the data that take place years in the future are due to definition changes or normalization that a forecaster would not have been attempting to forecast. As you can see in Figure A4, these changes can results in some large revisions to the PCE measure over time.

A2.1 Real-time forecasting procedure

The out-of-sample forecasts are performed recursively using the following procedure at each time period t:

Figure A4: Real-time data used for forecasting



1. Each candidate forecast is estimated on vintage τ data.

2. Each candidate forecast is used to construct a four-quarter-ahead forecast.

3. Equation (11) in Section 3.1.1 is estimated on the available real-time forecast errors for each of the candidate forecasts.

4. Equation (11) in Section 3.1.1 is used to predict the expected forecast errors of each model to construct $\hat{e}_{i,t+4}$ and $\hat{\xi}_{i,t+4}$.

5. The predicted errors are used to construct weights, and the weights are used to construct the combined forecast.

A2.2 Exploiting time-variation in Phillips curves

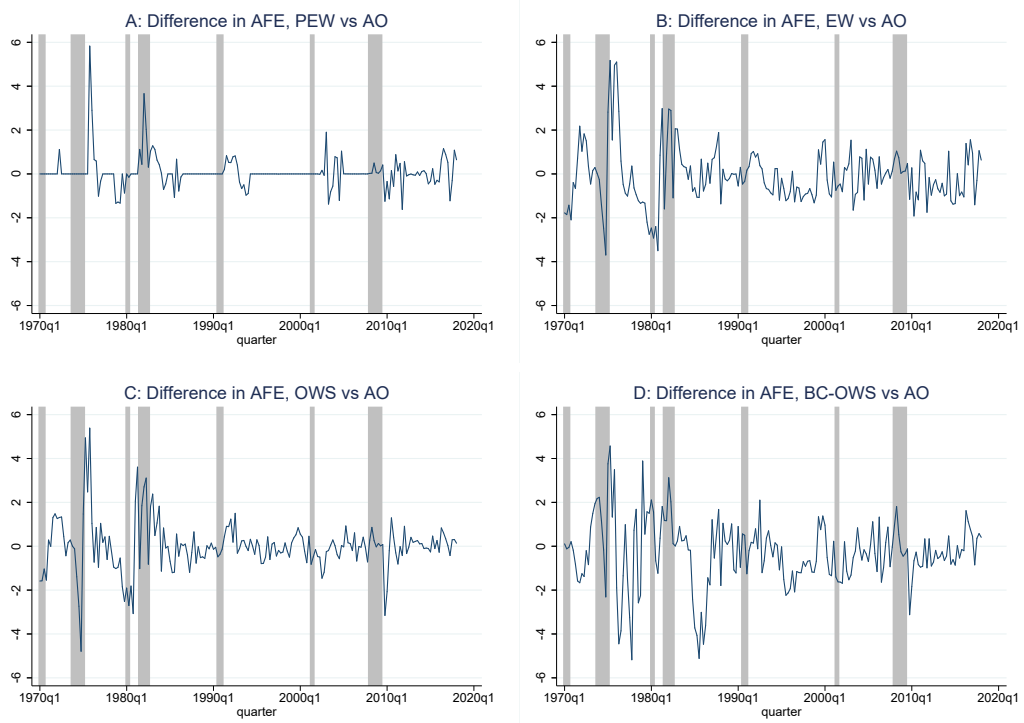
This section present supplementary results to complement the analysis in Section 3 on forecasting US inflation. The first set of results shows that the small full-sample gains observed for COW strategies actually coincide with economically meaningful periods. Specifically, most of the gains in forecast accuracy occur around recessions.

The second exercise reinforces the first by showing how weights shift in real time between Phillips curve like forecasts and the ARIMA forecasts.

A2.2.1 Economic significance

Although the average improvement in forecast accuracy of the COW strategies are not large in absolute terms, the timing of when the gains occur is economically significant. Figure A5 illustrates this point by comparing the observed difference in the absolute forecast errors (AFE) between the most accurate individual forecast, the AO forecast, to PEW, EW, OWS, and BC-OWS combined forecasts. Values that are above zero correspond to a more accurate combined forecast. Comparing PEW to the AO forecast, there are long periods where there is no difference. This result reflects the fact that the AO forecast is predicted to be the best forecast in many periods and given a weight near one. However, during and immediately

Figure A5: Economic significance



Notes: Shaded regions show the NBER recession dates. The lines show the difference in squared forecast errors for four-quarter-ahead forecasts of inflation between an AO model and a combined forecast. Values above zero correspond to a smaller squared forecast error in the quarter for the combined forecast.

following recessions, in real time, the PEW weights shift towards PC-type models, which generates large improvements in forecast accuracy. Recall that these are four-quarter-ahead forecasts. Therefore, by predicting the biases, we are able to somewhat capture in real time the time-varying efficiency of our different model specifications.

The EW and OWS strategies show similar improvements in forecast accuracy around recessions. However, these improvements are due to the hedging benefit of model averaging. Because forecast accuracy among the models diverges greatly around recessions, placing some weight on all models mitigates the impacts of the worst performing forecast. However, this comes at the expense of less weight on the best models. Therefore, although the overall forecast improvements of COW strategies are often small relative to other considered methods over the full sample, the improvements occur at relevant times – mainly around recessions – and are less prone to significant falls in accuracy overall relative to the naïve benchmark.

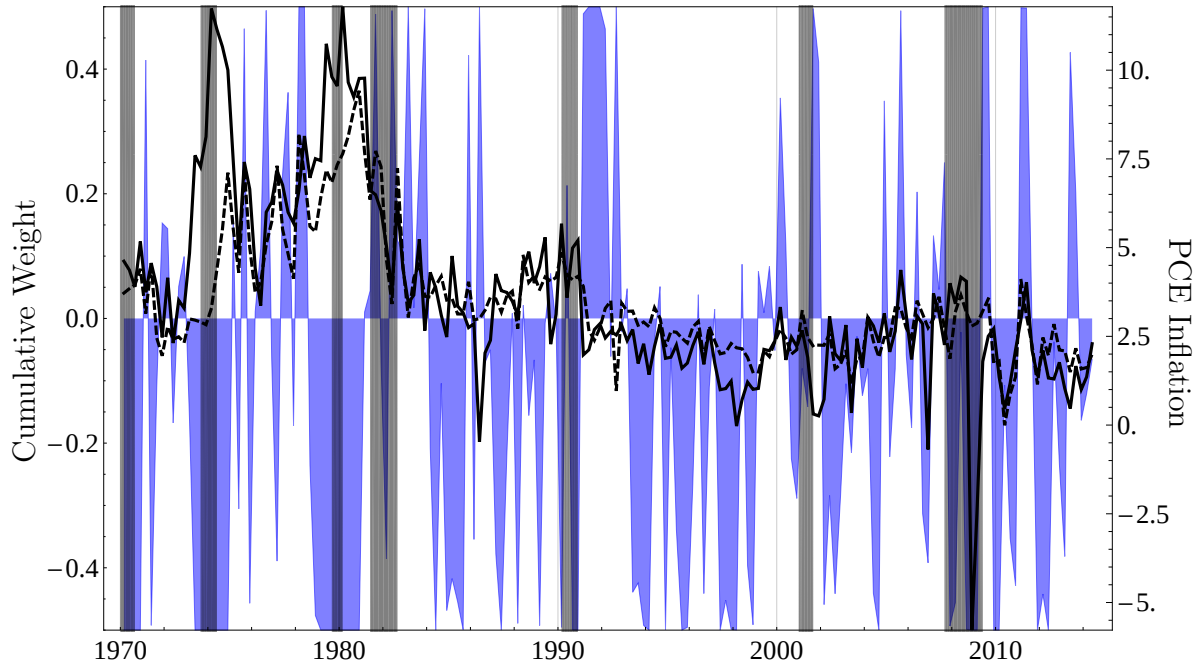


Figure A6: This figure depicts the PCE measure of US inflation from 1970Q1 to 2014Q1 (solid), the ex post predicted weights forecast (dashed), and the cumulative weight placed on the Phillips curve forecasts relative to equal weights for the ex post predicted weights (shaded blue). The dark bars indicate the NBER recession dates.

Finally, the BC-OWS case shows the perils of bias-correcting the underlying forecasts before combining. The large biases of the late 1970s and early 1980s are predicted to continue into the mid-1980s, which drives a significant and persistent deterioration in the combined forecast relative to the AO forecast. A similar event occurs in the mid-1990s. The COW strategies, on the other hand rely only on the relative size of the predicted bias to assign weights, which delivers more stable and accurate combined forecasts.

A2.2.2 Forward versus backward-looking weights and shifting weights around recessions

There is a known time variation in the forecast accuracy of Phillips curve specifications. This known time variation is one source of the predictable information in the forecast errors, which is exploited by our combination strategy. To illustrate how the forward-looking weights use this information, we compare a backward-looking strategy with a forward-looking strategy using six univariate and six Phillips curve forecasts of inflation. We use the Predicted

Exponential Weight given by Equation (12) and a backward-looking modification

$$\widehat{\boldsymbol{w}}^{*}_{\text{AO}}(I_T) = \frac{1}{\sum_{l=1}^{n} \exp(-\gamma \widetilde{b}^2_{l,T})} \left(\exp(-\gamma \widetilde{b}^2_{1,T}), \ldots, \exp(-\gamma \widetilde{b}^2_{n,T})\right)', \tag{A3}$$

where

$$\widetilde{b}_{i,T} = \frac{1}{4} \sum_{j=1}^{4} e_{i,T-j}.$$

We will refer to the backward-looking case as AO predicted weights because the prediction takes the same form as the AO forecast model used for inflation. The AO Predicted Weights are similar to the weights explored by Stock and Watson (2004) and capture the idea of weighting models by recent past performance. For Predicted Exponential Weights we use the output gap prediction. We set $\gamma = 5$ in both cases.

To assess how well the weights perform, we compare them against a counterfactual series of *ex post weights* that are constructed by using the realized *ex post* forecast error ($e_{i,T+4}$) of each model rather than $\widetilde{b}_{i,T}$ in the above weights equation. This reveals the maximum improvement in MSFE possible using this combination method. Figure A6 presents the ex post weights and their implied combined forecast for PCE inflation. The ex post weights produce an unbiased combined forecast that is a 26% improvement over both the AO and equal weights combined forecast in RMSFE.[1] The cumulative weights illustrated in the graph are constructed by summing the weights placed on the PC forecasts in each quarter and subtracting it from one half ($\sum_{i \in PC} w_{i,t} - 0.5$). Therefore, the graph provides an approximate description of how the cumulative weight on the PC forecasts shifts relative to equal weights over time. Points that are above zero indicate that greater than half of all weight is on the PC forecast specifications. Points below zero represent that greater than half of all weight is on the univariate forecast specifications.

Figure A7 shows the AO weights compared to the ex post weights. The backward-looking strategy results in a modest but statistically significant 5% loss in the relative RMSFE

[1]The AO and equal weights combined forecasts have a relative RMSFE of 1.0004. The benchmark combined weights also result in a slight increase in RMSFE compared to actually forecasting with the ex post best model in each period.

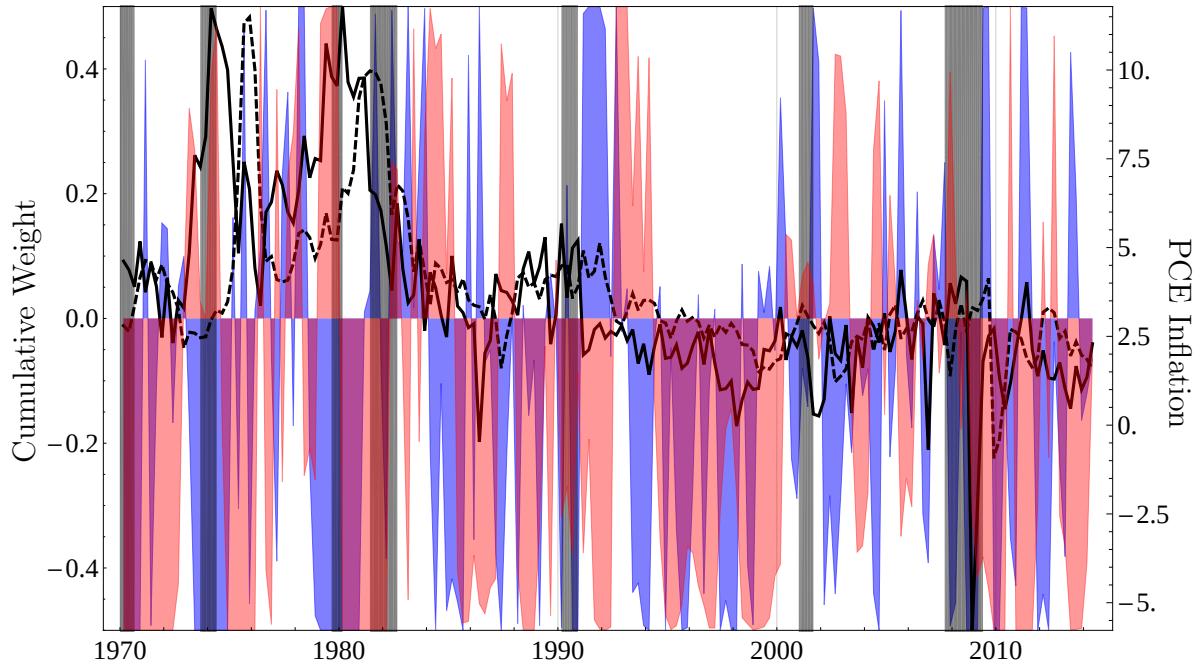


Figure A7: This figure depicts the PCE measure of US inflation from 1970Q1 to 2014Q1 (solid), the AO weights combined forecast (dashed), the cumulative ex post weights relative to equal weights (shaded blue), and the cumulative AO weights relative to equal weights (shaded red). The dark bars indicate the NBER recession dates.

compared to equal weights and the AO forecasts. The reason for why this strategy fails to improve upon equal weights is clearly visible in the figure. The backward-looking weights are negatively correlated with the ex post weights (correlation equal to -0.138). The strategy shifts weight to the PC forecasts after periods when the PC forecasts perform well, which is of course precisely when the strategy is about to lose forecast efficiency relative to the univariate forecasts.

Figures A6 and A7 also illustrate the relationship between the PC forecast efficiency and economic downturns. The ex post weights consistently shift toward the Phillips curve forecasts in the periods surrounding the NBER recession dates. A forward-looking strategy can take advantage of this regularity by shifting weight toward PC forecasts when real activity is weak and by shifting weights toward the univariate models when real activity is strong.

Figure A8 shows the weights for the forward-looking strategy plotted against the ex post weights. The Predicted Exponential Weights often deviate far from the ex post weights in

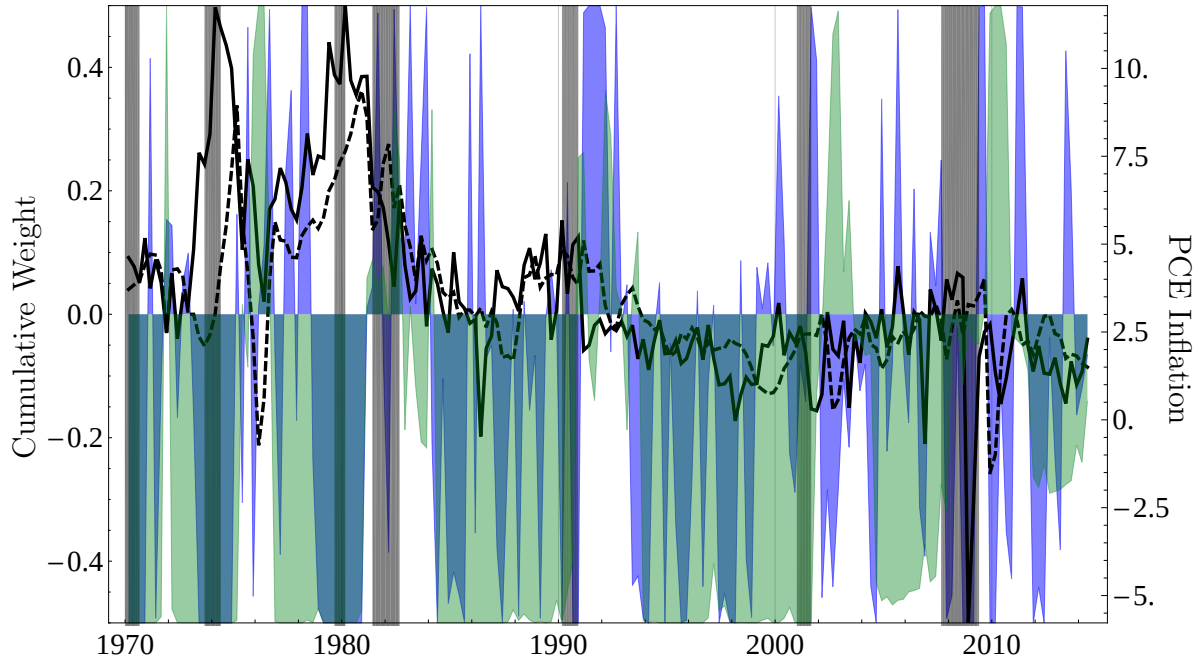


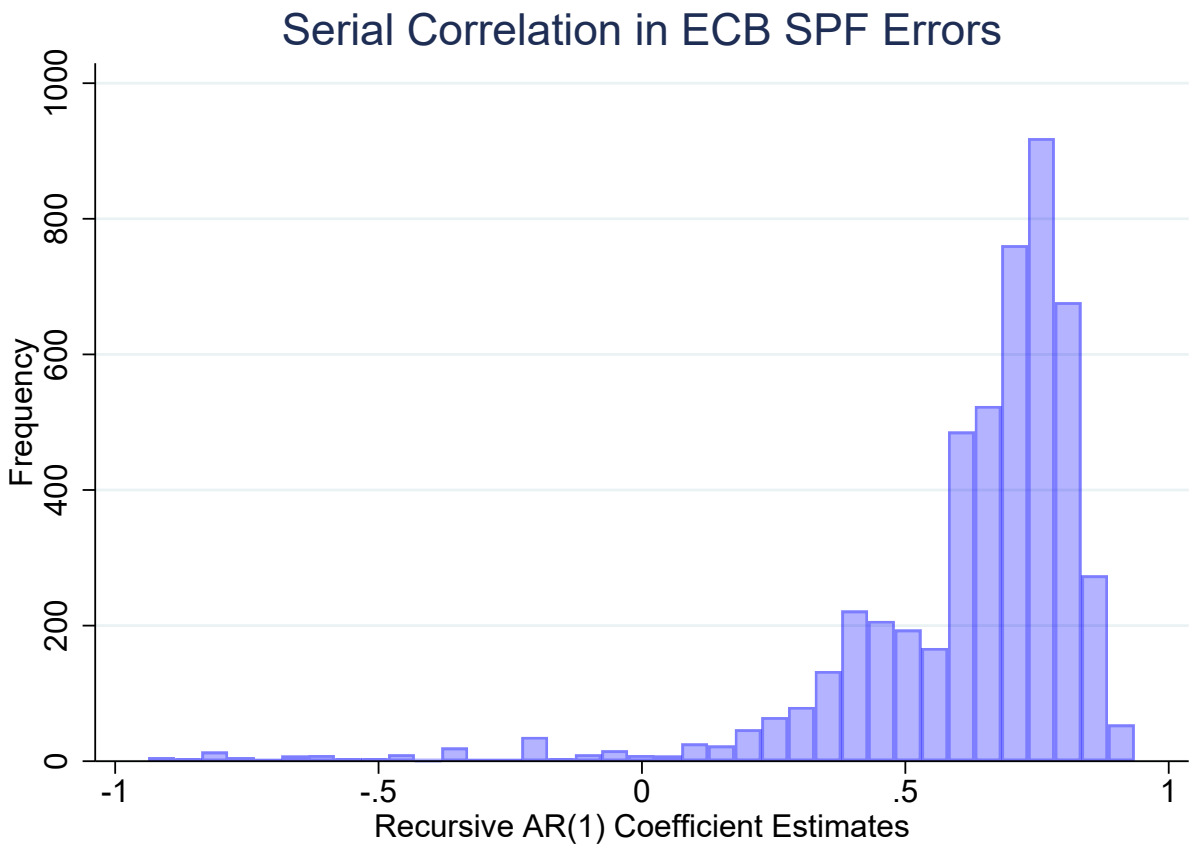
Figure A8: This figure shows the PCE measure of US inflation from 1970Q1 to 2014Q1 (solid), the Predicted Exponential Weights combined forecast (dashed), the cumulative ex post weights relative to equal weights (shaded blue), and the cumulative Predicted Exponential Weights relative to equal weights (shaded green). The dark bars indicate the NBER recession dates.

this case but are positively correlated with them over time (correlation equal to 0.178). The positive correlation translates into a statistically significant 7% improvement in RMSFE over both equal weights and the AO forecasts for the set of considered models. By forecasting the changes in the relative forecast accuracy of the models, the weights are able to shift in real time away from model specifications that are losing forecast accuracy to specifications that are gaining accuracy.

A3 Supplementary material for Section 4

We initiate our estimates of the bias for this exercise using the in-sample period 1999Q1 - 2001Q1. We estimate the AR(1) model using maximum likelihood and the Kalman filter recursively thereafter. The Kalman filter allows us to accommodate missing observations so that our bias predictions use all available information, subject to the real-time restrictions, to construct each bias forecast. Figure A9 shows a histogram of all the AR(1) coefficients that we recursively estimate for the individual survey participant's forecast errors. There is

Figure A9: AR(1) Coefficients for Conditional Bias Estimates from ECB SPF



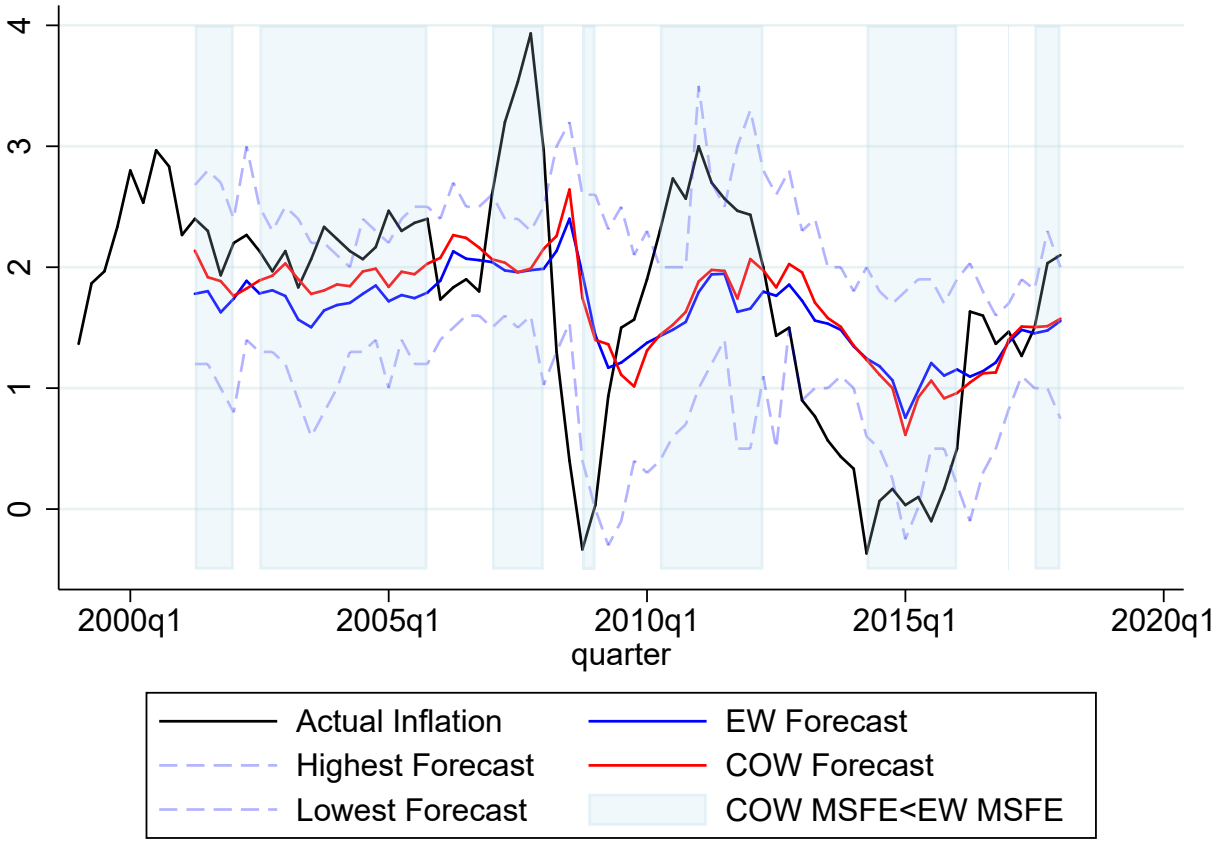
Notes: The figures shows a histogram of all the AR(1) coefficients recursively estimated for the individual survey participant's forecast errors in real-time.

considerable serial correlation that we can exploit in the survey to construct conditionally optimal weights.

Figure A10 shows the target inflation series, the range of the surveyed forecast responses in the ECB SPF for each quarter, the equal weights forecast, and the COW forecast. The shrinkage that is applied to the COW forecasts shrinks the weights toward equal weights. The shading indicates quarters where the COW forecast is more accurate and represents 68% of the out-of-sample period.

The principle combination strategy that we compare to COW is the combination approach of Issler and Lima (2009). Issler and Lima (2009) propose bias-corrected optimal weights and show that strategy yields the optimal combined forecast in an ergodic panel data setting. Like our setup, they assume that forecasters possess a menu of forecasts of $f_{i,t}^h$,

Figure A10: HCPI inflation and forecasts



Notes: The figures shows target real-time HCPI inflation series plotted against the highest and lowest forecasts recorded in the ECB Survey, the equal weights forecast, and the COW forecast. The shaded areas highlight quarter where the COW forecast is more accurate.

where $i = 1, 2, ..., N$ and h is the forecast horizon. The forecast errors of each forecast are assumed to be

$$(f_{i,t}^h - y_t) = k_i + \eta_t + \epsilon_{i,t},$$

where k_i is the fixed bias of each forecast, $\epsilon_{i,t}$ is the individual forecast error, η_t is an aggregate error common to all forecasts.

Issler and Lima show that given these assumptions an optimal forecasting strategy is to choose weights to minimize

$$E_t \left[\frac{1}{N} \sum_{i=1}^{N} \omega_{i,t} f_{i,t}^h - \frac{1}{N} \sum_{i=1}^{N} \omega_{i,t} \hat{k}_i - y_t \right]^2$$

where $\hat{k}_i$ is consistently estimated for the in-sample period ending in $t = R$ as

$$\hat{k}_i = \frac{1}{R} \sum_{t=1}^{R} f_{i,t}^h - \frac{1}{R} \sum_{t=1}^{R} y_t.$$

In the real-time forecasting exercises, we choose $w_{i,t} = 1/N$ and we update the estimates of k_i recursively as new data arrives.

A4 Supplementary material for Section 5

Figure A11 shows the individual forecasts and the real-time target series for each variable and country. The dashed line indicates the start of real-time forecasts. We make minimal adjustments to the real-time data set, which at times can exaggerate or ameliorate forecast biases as the sample sizes change.

The most significant biases are found in the direct forecasts (DF) of interest rates. The DF forecasts are constructed as

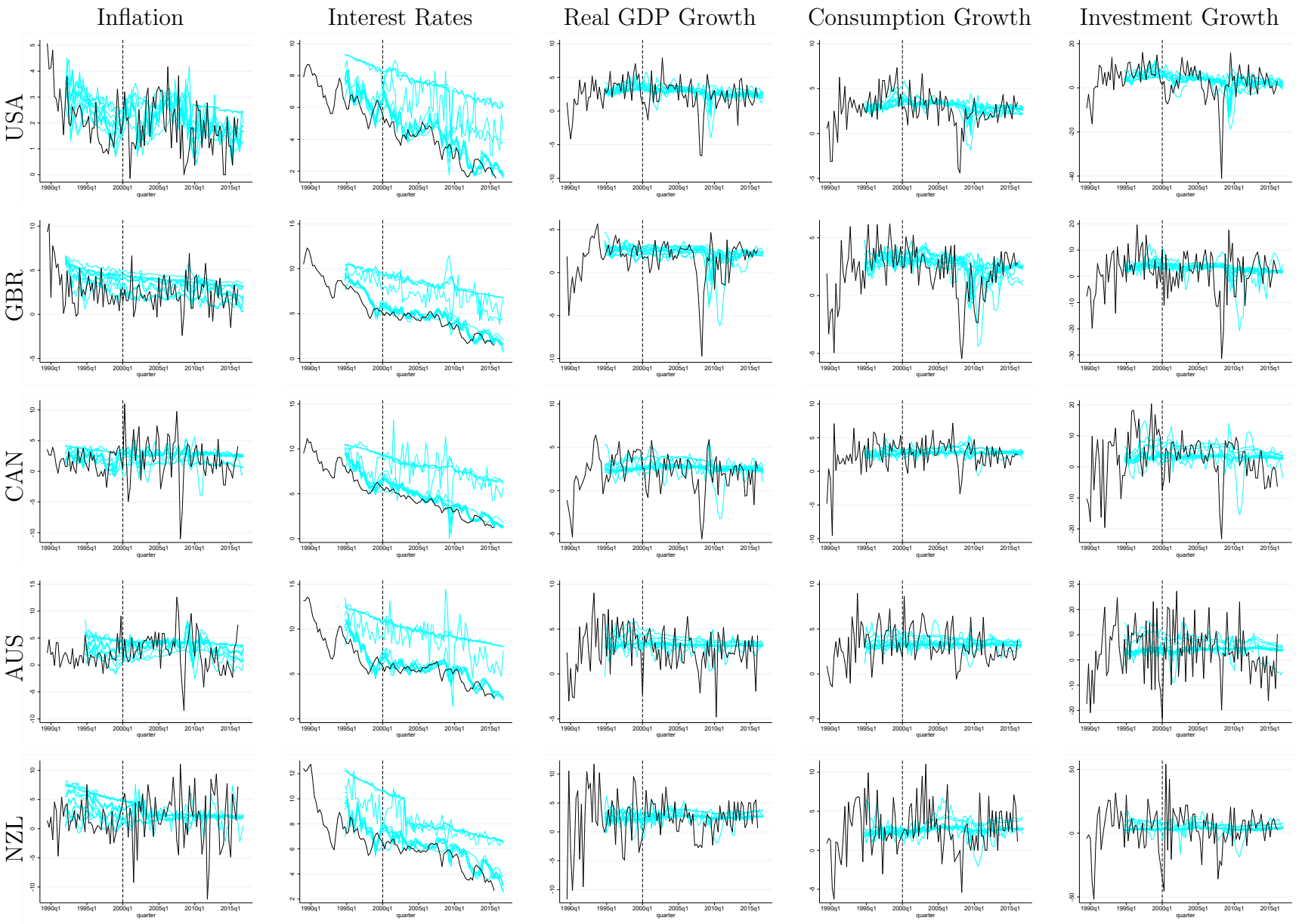
$$y_{j,t+4} = c + \beta x_{j,t} + \epsilon_{j,t+4}$$

where $x_{j,t}$ is real GDP growth when forecasting the interest rate. Real GDP growth does

not have a significant downward trend in any of the considered country to match the decline in the interest rate. Therefore, the level of the forecast is mostly determined by c, the intercept. This leads to the large and persistent biases seen in Figure A11 and documented in Table 4 in the main text, which reflects the fact that interest rates were high early in the sample. The effects of the high interest rates early in the sample is clearly seen in the NZL forecasts, where later vintages of real GDP have missing data pre-1980, which shorten the in-sample estimation period. We do not impute this data from earlier vintages, which means the average interest rate over the in-sample estimation period is lower for the forecasts made late in the sample.

The large biases are useful to make the point that bias correction may be a good strategy to pursue when there are obvious biases. We also do not want to alter the forecasts we have chosen ex post. We chose twelve models to forecast all five different variables of interest before knowing the outcome. In keeping with our real-time assumptions, ex ante, a forecaster never knows how accurate a forecast is. That is the reason why combination strategies are pursued.

Figure A11: Out-of-sample forecasts



Notes: Four quarter-ahead out-of-sample forecasts plotted against the target series. Real-time forecasts begin in 2000q1, which is indicated by the vertical dashed line.

References

Clark, T. E. and K. D. West (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of econometrics 138*(1), 291–311.

Elliott, G., A. Gargano, and A. Timmermann (2013). Complete subset regressions. *Journal of Econometrics 177*(2), 357–373.

Issler, J. V. and L. R. Lima (2009). A panel data approach to economic forecasting: The bias-corrected average forecast. *Journal of Econometrics 152*, 153—164.

Lima, L. R. and F. Meng (2017). Out-of-sample return predictability: A quantile combination approach. *Journal of Applied Econometrics 32*(4), 877–895.

Stock, J. H. and M. W. Watson (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting 23*(6), 405–430.