UNSW Business School Research Paper No. 2017 ECON 10

Conditionally Optimal Weights and Forward-Looking Approaches to Combining Forecasts

Christopher G. Gibbs
Andrey L. Vasnev

Conditionally Optimal Weights and Forward-Looking Approaches to Combining Forecasts

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

February 17, 2017

Abstract

In applied forecasting, there is a trade-off between in-sample fit and out-of-sample forecast accuracy. Parsimonious model specifications typically outperform richer model specifications. Consequently, there is often predictable information in forecast errors that is difficult to exploit. However, we show how this predictable information can be exploited in forecast combinations. In this case, optimal combination weights should minimize conditional mean squared error, or a conditional loss function, rather than the unconditional variance as in the commonly used framework of Bates and Granger (1969). We prove that our conditionally optimal weights lead to better forecast performance. The conditionally optimal weights support other forward-looking approaches to combining forecasts, where the forecast weights depend on the expected model performance. We show that forward-looking approaches can robustly outperform the random walk benchmark and many of the commonly used forecast combination strategies, including equal weights, in real-time out-of-sample forecasting exercises of inflation.

JEL Classifications: C18; C53; E31

Keywords: Forecast combination, conditionally optimal weights, forecast combination puzzle, inflation, Phillips curve

*Corresponding author contact information: School of Economics, UNSW, Sydney, christopher.gibbs@unsw.edu.au. We would like to thank Graham Elliott, George Evans, Jeremy Piger, Ken Wallis, James Morley, Benjamin Wong, Yinchu Zhu; conference participants of 32^{nd} Annual International Symposium of Forecasters, 2013 Canadian Economic Associations Meetings, the Continuing Education in Macroeconomics Workshop hosted by the University of Tasmania; seminar participants at Massey University, Auckland University of Technology, University of Sydney, and at 2016 SERG workshop; and the research staff at Reserve Bank of New Zealand for useful comments on earlier drafts of this paper. We would also like to thank the Reserve Bank of New Zealand for allowing use of their real-time dataset. The authors are responsible for any remaining errors.

1 Introduction

The standard approach in much of the literature on forecast combination is to construct the combination weights based on the past performances of the individual forecasts.[1] This backward-looking approach is very sensible, but it also has generated a puzzle. The forecast combination puzzle is the empirical finding that simple forecast combination strategies, such as equal weights (averaging), produce the most reliably accurate forecasts. This is despite the fact that the equal weights strategy is only optimal under very restrictive assumptions on the correlation and covariance of the individual models' forecast errors and despite the fact that significant past differences often exist in forecast accuracy that should be exploitable to create accurate combined forecasts.[2] The puzzle, therefore, is that most forecast combination strategies work in theory but not in practice when compared to equal weights.[3]

In this paper, we show that a key contributor to the poor performance of optimal forecast combination strategies in practice is that the underlying forecasts typically have serially correlated and predictable errors. We show that when forecast errors are predictable, the optimal combination weights are those that minimize the mean squared error conditional on the predictable information. We prove that our conditionally optimal weights lead to better forecast performance under a general loss function. The theoretical results suggest that dynamic forecast combination strategies should be forward-looking in the sense that weights should be assigned based on the *expected* forecast performance rather than the past forecast performance.

Building on our theoretical results, we propose a number of operational forward-

[1]See Timmermann (2006) for a survey of the literature and Elliott and Timmermann (2016) for the most recent rigorous treatment of the relevant issues.

[2]An empirical example of the forecast combination puzzle for inflation is presented in Stock and Watson (2003). To our knowledge, the first formal reference to the forecast combination puzzle in the literature is Stock and Watson (2004); however, the results are certainly known in the literature, at least dating back to Bates and Granger (1969).

[3]Surveys and comments on the literature are found in Clemen (1989), Granger (1989), Diebold and Lopez (1996), Timmermann (2006), and Wallis (2011).

looking combination strategies that construct combination weights by predicting the forecast error of each model out to a forecast horizon of interest and assign weights according to each model's relative expected squared error. We use the term "operational" to describe our proposed strategies because in practice the true conditional optimal weights that we derive still suffer from small sample estimation problems that plague all optimal forecast combination strategies. Therefore, we take the basic principle from our conditionally optimal weights to construct simple and transparent combined forecasts as a proof-of-concept for improvements that may be obtained through forward-looking combination strategies relative to equal weights.

In practice, our operational strategies are feasible because there is almost always predictable information in actual forecast errors due to parsimonious specification choices to avoid data overfitting or simply because real-world data often contain structural breaks. The information in the errors is typically not exploitable by correcting individual model specifications precisely because of overfitting issues. Such information is, however, exploitable for constructing combined forecasts. We show that forward-looking combination strategies can significantly and robustly outperform individual time-series models (including the random walk model), equal weights combined forecasts, and other commonly used forecast combination strategies in real-time out-of-sample forecasting experiments of US and New Zealand inflation. We select inflation as the empirical application because 1) it is very difficult to forecast, see Stock and Watson (2007), and 2) because it is well documented that time variation exists in the forecast accuracy of Phillips curve model specifications, see Stock and Watson (2009), which our strategy is well suited to exploit.

We also address an external validity concern that exists when comparing forecast combination strategies on fixed sets of forecasting models to demonstrate that our preferred operational strategy robustly outperforms the equal weights strategy. The individual forecasting specifications chosen by a researcher dictate the improvements in forecast accuracy that are possible, and these choices drive the reported results. An example of this is when a number of poor performing forecasts are considered among the set of forecasts. Sophisticated forecast combination strategies often easily detect the poor forecasts and will outperform equal weights. However, the forecast combination puzzle reemerges if the set of models is trimmed to include only the best performing models. This concern is addressed by conducting a forecasting tournament that compares multiple forecast combination strategies on all possible subsets of the best models considered in this paper to show that the performance of forward-looking weights is robust to model choices. In addition, the fact that we obtain similar results for both the US and New Zealand data suggests that the information that we exploit for forecast combination is not country specific.

The concept of conditionally optimal combinations is approached by Aiolfi and Timmermann (2006). They note that there is persistence in relative forecast performance among linear and nonlinear time-series models used to forecast a wide range of macroeconomic variables and that this persistence can be exploited to select forecasts and construct weights. However, they limit their approach to the weights estimated by least squares regression of the actual realization on the individual forecasts. Thus, Aiolfi and Timmermann (2006) restrict the information set used for conditioning and return to a backward-looking framework in that their conditional combination strategies are based on recent historical forecasting performances. We provide a general framework that allows conditioning on any available information set and that encompasses Aiolfi and Timmermann's weights. More importantly, all of our operational strategies are forward-looking and based on predictions of future forecast errors rather than focusing solely on relative past performance. We also prove that our conditionally optimal weights lead to better forecast performance in terms of expected loss relative to unconditional weights and that extending the information used for conditioning further improves the results.

There are two related papers that support our forward-looking approach. Giacomini and Rossi (2009) show that detecting changes in relative forecast accuracy is possible in real time and Timmermann and Zhu (2016) show that a forward-looking approach to model selection is useful in situations with weak predictors due to estimation error. As in those papers, we show that changes in accuracy of Phillips curve forecasts are predictable in real time. We, however, use the predictions to construct combined forecasts rather than for model selection.

Our approach is also related to Wallis and Whitley (1991) and Clements and Hendry (1996), who study the use of forecast errors to improve forecast efficiency through a strategy known as intercept correction. Intercept correction uses the most recently observed forecast errors to correct the bias of a point forecast by adding the errors to the next forecast to set the model back on track. Wallis and Whitley (1991) finds that intercept correction produces modest improvements over an uncorrected model for forecasts of UK inflation and other macroeconomic variables. We explore intercept correction relative to conditional optimal forecast combinations in our section on robustness.

The remainder of the paper is organized as follows. In Section 2 we provide the theory behind conditionally optimal weights and discuss forward-looking strategies. Section 3 presents our application, where we use conditionally optimal weights and forward-looking strategies to forecast US data in real time. Section 4 provides further analysis and robustness of our proposed weights including the forecast tournament and real-time forecasts of New Zealand inflation. Section 5 concludes this work.

2 Conditionally Optimal Weights

There are two explanations for why the forecast combination puzzle exists. The first explains the failure of optimal weighting strategies in small sample environments. Optimal weighting strategies require the estimation of combination weights from the joint distribution of all the considered models' forecast errors. This introduces significant estimation uncertainty because the number of models considered is typically large relative to the amount of available data. Smith and Wallis (2009) show using Monte Carlo experiments that equal weights generally provide more accurate forecasts compared to estimated optimal weights, particularly when the optimal weights are close to being equal. Claeskens et al. (2016) formalize this argument and provide a proof showing how the estimation of the weights influences the results. Therefore, the literature often recommends employing combination strategies that do not require estimation of the weights.

The second explanation is that economic forecasting models are often misspecified and many economic data series are subject to relatively frequent structural breaks. Hendry and Clements (2004) show that the equal weights strategy is an effective strategy in this case because it mitigates the different biases that arise in differently specified models. In particular, differently specified models may manifest bias in opposite directions following a change in the data-generating process, which is averaged out in the combined forecast.

The second explanation, however, implies that in practice there is predictable information that is not taken into account by classical optimal weights. The underlying forecasts' errors exhibit serial correlation due to structural breaks and other time-varying misspecifications of the forecasting models. Therefore, the optimal forecast combination weights should condition on this predictable information, which implies minimizing the conditional mean squared error (MSE),

$$\text{MSE} = \text{Bias}^2 + \text{var},$$

rather than the unconditional variance.

2.1 The case of two forecasts

To illustrate this point, assume that we need to forecast y_{T+1} and that we have two individual forecasts available, $f_{1,T+1}$ and $f_{2,T+1}$, which can be combined linearly

$$f_{c,T+1} = w f_{1,T+1} + (1-w) f_{2,T+1}.$$

In the classical framework of Bates and Granger (1969), if the individual forecasts are unbiased, i.e., the forecast errors $e_{1,T+1} = y_{T+1} - f_{1,T+1}$ and $e_{2,T+1} = y_{T+1} - f_{2,T+1}$ have zero expectation, then the error of the combined forecast $e_{c,T+1} = y_{T+1} - f_{c,T+1} = we_{1,T+1} + (1-w)e_{2,T+1}$ will have zero expectation, and its variance is

$$\text{var}(e_{c,T+1}) = w^2\sigma_{e_1}^2 + (1-w)^2\sigma_{e_2}^2 + 2w(1-w)\rho_{e_1,e_2}\sigma_{e_1}\sigma_{e_2},$$

where $\sigma_{e_1}^2 = \text{var}(e_{1,T+1})$, $\sigma_{e_2}^2 = \text{var}(e_{2,T+1})$ and $\rho_{e_1,e_2} = \text{corr}(e_{1,T+1}, e_{2,T+1})$. The variance of the combined forecast is minimized when

$$w^* = \frac{\sigma_{e_2}^2 - \rho_{e_1,e_2}\sigma_{e_1}\sigma_{e_2}}{\sigma_{e_1}^2 + \sigma_{e_2}^2 - 2\rho_{e_1,e_2}\sigma_{e_1}\sigma_{e_2}}. \tag{1}$$

We will refer to w^* as *the unconditionally optimal weight*.

To extend the classical framework, assume that an information set I_T is available. We can apply the same reasoning as before but conditionally on I_T. The errors can be decomposed as

$$e_{1,T+1} = b_{1,T} + \xi_{1,T+1}$$

and

$$e_{2,T+1} = b_{2,T} + \xi_{2,T+1},$$

where $b_{1,T} = \text{E}(e_{1,T+1}|I_T)$, $b_{2,T} = \text{E}(e_{2,T+1}|I_T)$, and $\text{E}(\xi_{1,T+1}|I_T) = \text{E}(\xi_{2,T+1}|I_T) = 0$. Note that there is no contradiction with the unbiasedness of the original forecasts as long as unconditionally $\mathrm{E}(b_{1,T}) = \mathrm{E}(b_{2,T}) = 0$. A simple AR(1) model can capture the autocorrelation of the forecast errors if $b_{1,T} = \gamma_1 e_{1,T}$ and $b_{2,T} = \gamma_2 e_{2,T}$, but I_T can potentially have additional variables for error prediction.

Conditionally on I_T, the error of the combined forecast will have non-zero expectation

$$\mathrm{E}(e_{c,T+1}|I_T) = wb_{1,T} + (1-w)b_{2,T}$$

and its variance

$$\mathrm{var}(e_{c,T+1}|I_T) = w^2\sigma^2_{\xi_1} + (1-w)^2\sigma^2_{\xi_2} + 2w(1-w)\rho_{\xi_1,\xi_2}\sigma_{\xi_1}\sigma_{\xi_2},$$

where $\sigma^2_{\xi_1} = \mathrm{var}(\xi_{1,T+1}|I_T)$, $\sigma^2_{\xi_2} = \mathrm{var}(\xi_{2,T+1}|I_T)$, and $\rho_{\xi_1,\xi_2} = \mathrm{corr}(\xi_{1,T+1}, \xi_{2,T+1}|I_T)$. In this situation, one should consider minimizing the conditional mean squared error to balance the bias and the variance components simultaneously

$$\mathrm{MSE}(w) = (wb_{1,T} + (1-w)b_{2,T})^2 + w^2\sigma^2_{\xi_1} + (1-w)^2\sigma^2_{\xi_2} + 2w(1-w)\rho_{\xi_1,\xi_2}\sigma_{\xi_1}\sigma_{\xi_2}$$

and the optimal solution in this case is

$$w^*(I_T) = \frac{\sigma^2_{\xi_2} + b^2_{2,T} - \rho_{\xi_1,\xi_2}\sigma_{\xi_1}\sigma_{\xi_2} - b_{1,T}b_{2,T}}{\sigma^2_{\xi_1} + b^2_{1,T} + \sigma^2_{\xi_2} + b^2_{2,T} - 2\rho_{\xi_1,\xi_2}\sigma_{\xi_1}\sigma_{\xi_2} - 2b_{1,T}b_{2,T}}. \qquad (2)$$

We will refer to $w^*(I_T)$ as *the conditionally optimal weight* and I_T explicitly indicates the information set used for conditioning.

There are several special cases that help to understand the behavior of the conditionally optimal weight.

1. If $b_{1,T} = b_{2,T}$ then the biases cancel out and are not used to determine the optimal

weight. The optimal conditional solution (2) simplifies to

$$\frac{\sigma_{\xi_2}^2 - \rho_{\xi_1,\xi_2}\sigma_{\xi_1}\sigma_{\xi_2}}{\sigma_{\xi_1}^2 + \sigma_{\xi_2}^2 - 2\rho_{\xi_1,\xi_2}\sigma_{\xi_1}\sigma_{\xi_2}}$$

which is similar to the classical (unconditional) solution (1) that is well studied.

2. If $\rho_{\xi_1,\xi_2} = 0$, $\sigma_{\xi_1} = \sigma_{\xi_2}$ and $b_{1,T} = b_{2,T}$ then the conditionally optimal weight $w^*(I_T)$ is equal to $1/2$, which is the same as the classical (unconditional) solution w^* given by (1) when $\rho_{e_1,e_2} = 0$ and $\sigma_{e_1} = \sigma_{e_2}$.

3. If $\rho_{\xi_1,\xi_2} = 0$ and $b_{2,T} = 0$, then the conditionally optimal solution is

$$\frac{\sigma_{\xi_2}^2}{\sigma_{\xi_1}^2 + b_{1,T}^2 + \sigma_{\xi_2}^2}.$$

In the classical (unconditional) formula (1), the ratio between variances determines the weight if the correlation is zero (i.e., $w^* = \frac{\sigma_{e_2}^2}{\sigma_{e_1}^2+\sigma_{e_2}^2}$); the conditional bias now also plays a similar role. The larger $b_{1,T}^2$ is, the smaller is the weight $w^*(I_T)$. If $\sigma_{\xi_1} = \sigma_{\xi_2}$, then the weights are not equal, which is different from the classical solution of $1/2$ given by (1) when $\rho_{e_1,e_2} = 0$ and $\sigma_{e_1} = \sigma_{e_2}$.

4. If $\rho_{\xi_1,\xi_2} = 0$, then

$$w^*(I_T) = \frac{\sigma_{\xi_2}^2 + b_{2,T}^2 - b_{1,T}b_{2,T}}{\sigma_{\xi_1}^2 + b_{1,T}^2 + \sigma_{\xi_2}^2 + b_{2,T}^2 - 2b_{1,T}b_{2,T}}$$

and if $b_{1,T}b_{2,T} < 0$, i.e., the biases work in different directions, then the optimal weight is non-negative $w^*(I_T) > 0$. However, if $b_{1,T}b_{2,T} > 0$, then the optimal weight can be negative. In other words, if we know that both forecasts are expected to over (under) estimate, then we need to select the combination that is smaller (larger) than both forecasts.

2.2 General case

The conditional approach can easily be adapted for a general case with n forecasts and different forecast horizons. Assume that we need to forecast $y_{T+h} \in \mathbb{R}$ and consider the vector of h-step-ahead forecasts

$$\boldsymbol{f}_{T+h} = (f_{1,T+h}, f_{2,T+h}, \ldots, f_{n,T+h})' \in \mathbb{R}^k$$

and the information set I_T available to us at time T. The forecast vector $\boldsymbol{f}_{T+h}$ needs to be mapped to the real line, and following Aiolfi and Timmermann (2006), we limit the analysis to a linear combination that involves selecting a vector of weights

$$\boldsymbol{w} = (w_1, w_2, \ldots, w_n)'$$

to produce the combined forecast $f_{c,T+h} = \boldsymbol{w'f}_{T+h}$. We denote the vector of forecasting errors as

$$\boldsymbol{e}_{T+h} = y_{T+h}\boldsymbol{\iota} - \boldsymbol{f}_{T+h},$$

where $\boldsymbol{\iota}$ is a vector of ones, and the error of the combined forecast as

$$e_{c,T+h} = y_{T+h} - f_{c,T+h} = \boldsymbol{w'e}_{T+h}.$$

Assuming that the loss function $L(\cdot)$ that describes how costly it is to use an imperfect forecast depends only on the forecast error, $e_{c,T+h}$, the conditionally optimal combination weights, $\boldsymbol{w}^*(I_T)$, solves the problem

$$\boldsymbol{w}^*(I_T) = \arg\min_{\boldsymbol{w}} \mathrm{E}[L(e_{c,T+h})|I_T]. \tag{3}$$

Under mean squared error (MSE) loss, $L(e) = e^2$, only the first two conditional moments influence the optimal weight, and the optimization problem can be solved explicitly. We also assume that the forecasts are unbiased, $\mathrm{E}(\boldsymbol{e}_{T+h}) = \mathbf{0}$; thus we solve the optimization problem (3) subject to the restriction that the weights sum up to one[4], $\boldsymbol{w}'\boldsymbol{\iota} = 1$.

If the errors of the original forecasts are decomposed into two parts

$$\boldsymbol{e}_{T+h} = \boldsymbol{b}_T + \boldsymbol{\xi}_{T+h},$$

where $\boldsymbol{b}_T = \mathrm{E}(\boldsymbol{e}_{T+h}|I_T)$ and $\mathrm{E}(\boldsymbol{\xi}_{T+h}|I_T) = 0$, then

$$\mathrm{MSE}(\boldsymbol{w}) = (\boldsymbol{w}'\boldsymbol{b}_T)^2 + \boldsymbol{w}'\Sigma_\xi\boldsymbol{w} = \boldsymbol{w}'(\Sigma_\xi + \boldsymbol{b}_T\boldsymbol{b}_T')\boldsymbol{w},$$

where $\Sigma_\xi = \mathrm{var}(\boldsymbol{\xi}_{T+h}|I_T) = \mathrm{E}(\boldsymbol{\xi}_{T+h}\boldsymbol{\xi}_{T+h}'|I_T)$, and it is minimized by *the conditionally optimal weights*

$$\boldsymbol{w}^*(I_T) = \frac{[\Sigma_\xi + \boldsymbol{b}_T\boldsymbol{b}_T']^{-1}\boldsymbol{\iota}}{\boldsymbol{\iota}'[\Sigma_\xi + \boldsymbol{b}_T\boldsymbol{b}_T']^{-1}\boldsymbol{\iota}}. \tag{4}$$

The minimum MSE that is achieved by the conditionally optimal weights (4) is

$$\mathrm{MSE}(\boldsymbol{w}^*(I_T)) = \frac{1}{\boldsymbol{\iota}'[\Sigma_\xi + \boldsymbol{b}_T\boldsymbol{b}_T']^{-1}\boldsymbol{\iota}}. \tag{5}$$

Naturally, $\boldsymbol{b}_T$, Σ_ξ, $\mathrm{MSE}(\boldsymbol{w})$ and the optimal solution depend on I_T, but to keep the notation simple we specify this dependency explicitly only for the optimal solution $\boldsymbol{w}^*(I_T)$. This highlights the fact that the conditionally optimal weights $\boldsymbol{w}^*(I_T)$ are time varying because they depend on the information I_T available at time T. Without loss of generality, we assume Σ_ξ to be constant for the reminder of this section.[5]

For comparison, the well-established classical results, see Elliott (2011), for *the un-*

[4]The case of biased forecasts can be addressed by including a constant in $\boldsymbol{f}_{T+h}$ similar to Aiolfi and Timmermann (2006).

[5]The adaptation of our theory to cover the case of time-varying Σ_ξ (e.g., when it follows a GARCH model) is straightforward.

conditionally optimal weights

$$\boldsymbol{w}^* = \frac{\Sigma_e^{-1}\boldsymbol{\iota}}{\boldsymbol{\iota}'\Sigma_e^{-1}\boldsymbol{\iota}} \tag{6}$$

are based on the unconditional variance of the errors $\Sigma_e = \mathrm{var}(\boldsymbol{e}_{T+1}) = \Sigma_\xi + \mathrm{E}(\boldsymbol{b}_T\boldsymbol{b}_T')$ and the minimum of $\mathrm{var}(e_{c,T+h}) = w'\Sigma_e w$ achieved by $\boldsymbol{w}^*$ is

$$\frac{1}{\boldsymbol{\iota}'\Sigma_e^{-1}\boldsymbol{\iota}}. \tag{7}$$

There are several interesting observations that help us understand the new conditionally optimal weights.

1. Using unconditional weights $\boldsymbol{w}^*$ is equivalent to using *no information* to predict the errors, i.e., $\boldsymbol{w}^* = \boldsymbol{w}^*(\emptyset)$.

2. If $\boldsymbol{b}_T$ is proportional to $\boldsymbol{\iota}$, i.e., the predictable parts are the same for all forecasts, then $\boldsymbol{b}_T$ does not play a role in the conditionally optimal weight, i.e.,

$$\boldsymbol{w}^*(I_T) = \boldsymbol{w}^\dagger = \frac{\Sigma_\xi^{-1}\boldsymbol{\iota}}{\boldsymbol{\iota}'\Sigma_\xi^{-1}\boldsymbol{\iota}}.$$

3. Without predictability, i.e., if $\boldsymbol{b}_T = \mathbf{0}$, we have $\boldsymbol{w}^*(I_T) = \boldsymbol{w}^*$ because of $\Sigma_\xi = \Sigma_e$ in this case.

4. The presence of bias, i.e., if $\boldsymbol{b}_T \neq \mathbf{0}$, will increase the MSE:

$$\frac{1}{\boldsymbol{\iota}'\Sigma_\xi^{-1}\boldsymbol{\iota}} < \frac{1}{\boldsymbol{\iota}'[\Sigma_\xi + \boldsymbol{b}_T\boldsymbol{b}_T']^{-1}\boldsymbol{\iota}}.$$

5. If $\boldsymbol{w}^*$ is used rather than $\boldsymbol{w}^*(I_T)$ but $\boldsymbol{b}_T \neq \mathbf{0}$, then the minimum of the MSE given by (5) will not be achieved, i.e.,

$$\mathrm{MSE}(\boldsymbol{w}^*(I_T)) = \frac{1}{\boldsymbol{\iota}'[\Sigma_\xi + \boldsymbol{b}_T\boldsymbol{b}_T']^{-1}\boldsymbol{\iota}} < \boldsymbol{w}^{*'}[\Sigma_\xi + \boldsymbol{b}_T\boldsymbol{b}_T']\boldsymbol{w}^* = \frac{\boldsymbol{\iota}'\Sigma_e^{-1}[\Sigma_\xi + \boldsymbol{b}_T\boldsymbol{b}_T']\Sigma_\mathrm{e}^{-1}\boldsymbol{\iota}}{[\boldsymbol{\iota}'\Sigma_e^{-1}\boldsymbol{\iota}]^2}.$$

6. If the bias is ignored, i.e., $\boldsymbol{w}^\dagger = \frac{\Sigma_\xi^{-1}\boldsymbol{\iota}}{\boldsymbol{\iota}'\Sigma_\xi^{-1}\boldsymbol{\iota}}$ is used rather than $\boldsymbol{w}^*(I_T)$ but $\boldsymbol{b}_T \neq \boldsymbol{0}$, then the minimum of the MSE given by (5) will not be achieved, i.e.,

$$\mathrm{MSE}(\boldsymbol{w}^*(I_T)) = \frac{1}{\boldsymbol{\iota}'[\Sigma_\xi + \boldsymbol{b}_T\boldsymbol{b}_T']^{-1}\boldsymbol{\iota}} < \boldsymbol{w}^{\dagger\prime}[\Sigma_\xi + \boldsymbol{b}_T\boldsymbol{b}_T']\boldsymbol{w}^\dagger = \frac{1}{\boldsymbol{\iota}'\Sigma_\xi^{-1}\boldsymbol{\iota}} + \left[\frac{\boldsymbol{\iota}'\Sigma_\xi^{-1}\boldsymbol{b}_T}{\boldsymbol{\iota}'\Sigma_\xi^{-1}\boldsymbol{\iota}}\right]^2$$

Using the Cauchy-Schwarz inequality,

$$(\boldsymbol{\iota}'\Sigma_\xi^{-1}\boldsymbol{b}_T)^2 \leq (\boldsymbol{\iota}'\Sigma_\xi^{-1}\boldsymbol{\iota})(\boldsymbol{b}_T'\Sigma_\xi^{-1}\boldsymbol{b}_T)$$

we have another upper bound,

$$\mathrm{MSE}(\boldsymbol{w}^*(I_T)) < \frac{1}{\boldsymbol{\iota}'\Sigma_\xi^{-1}\boldsymbol{\iota}}\left[1 + \boldsymbol{b}_T'\Sigma_\xi^{-1}\boldsymbol{b}_T\right].$$

7. Using the Sherman-Morrison formula

$$[\Sigma_\xi + \boldsymbol{b}_T\boldsymbol{b}_T']^{-1} = \Sigma_\xi^{-1} - \frac{\Sigma_\xi^{-1}\boldsymbol{b}_T\boldsymbol{b}_T'\Sigma_\xi^{-1}}{1 + \boldsymbol{b}_T'\Sigma_\xi^{-1}\boldsymbol{b}_T}$$

it is possible to rewrite $\mathrm{MSE}(\boldsymbol{w}^*(I_T))$,

$$\frac{1}{\boldsymbol{\iota}'[\Sigma_\xi + \boldsymbol{b}_T\boldsymbol{b}_T']^{-1}\boldsymbol{\iota}} = \frac{1}{\boldsymbol{\iota}'\Sigma_\xi^{-1}\boldsymbol{\iota} - \frac{(\boldsymbol{\iota}'\Sigma_\xi^{-1}\boldsymbol{b}_T)^2}{1+\boldsymbol{b}_T'\Sigma_\xi^{-1}\boldsymbol{b}_T}}$$

which, in combination with the Cauchy-Schwarz inequality, provides us with the same upper bound,

$$\mathrm{MSE}(\boldsymbol{w}^*(I_T)) \leq \frac{1}{\boldsymbol{\iota}'\Sigma_\xi^{-1}\boldsymbol{\iota}}\left[1 + \boldsymbol{b}_T'\Sigma_\xi^{-1}\boldsymbol{b}_T\right].$$

8. The effect of the bias on the MSE of the combination can be investigated via the

differential

$$d\,\mathrm{MSE}(\boldsymbol{w}^*(I_T)) = \frac{\boldsymbol{\iota}'[\Sigma_\xi + \boldsymbol{b}_T\boldsymbol{b}_T']^{-1}\{(db)b' + b(db)'\}[\Sigma_\xi + \boldsymbol{b}_T\boldsymbol{b}_T']^{-1}\boldsymbol{\iota}}{(\boldsymbol{\iota}'[\Sigma_\xi + \boldsymbol{b}_T\boldsymbol{b}_T']^{-1}\boldsymbol{\iota})^2}.$$

To conclude the theoretical part we formulate the following central result that formalizes a very intuitive idea that using more information allows us to construct a better combination.

Theorem 1 *Given that the first and the second conditional and unconditional moments exist, the following inequalities hold:*

(a) for conditional and unconditional MSE,

$$\mathrm{E}(\mathrm{MSE}(\boldsymbol{w}^*(I_T))) \leq \mathrm{MSE}(w^*),$$

or equivalently

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

or equivalently

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

(b) for conditional MSE when two information sets $J_T \subset I_T$ are available,

$$\mathrm{E}(\mathrm{MSE}(\boldsymbol{w}^*(I_T))|J_T) \leq \mathrm{MSE}(\boldsymbol{w}^*(J_T)),$$

or equivalently

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

or equivalently

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

(c) for a convex loss function $L(\cdot)$ and $J_T \subset I_T$,

$$\mathrm{E}\left[\left.\min_{\boldsymbol{w}} \mathrm{E}[L(e_{c,T+h})|I_T]\right| J_T\right] \leq \min_{\boldsymbol{w}} \mathrm{E}[L(e_{c,T+h})|J_T]$$

if the conditional expectations and the solutions of the minimization problems exist.

Proof See Appendix B.

Theorem 1 has three parts, and the first two parts have three formulations. Part (a) compares the conditional and unconditional approaches, whereas part (b) presents a more general result when there are two information sets I_T and J_T available and I_T is an extension of J_T, i.e., it contains more information. Part (a) can be viewed as a special case of part (b) when $J_T = \emptyset$. Part (c) covers the case of a general loss function.

The first formulation in parts (a) and (b) is in terms of the precision achieved by the combinations as measured by the MSE. In part (a), to make the conditional and unconditional MSE comparable, we need to examine $\mathrm{E}(\mathrm{MSE}(\boldsymbol{w}^*(I_T)))$. In part (b), we need to condition on the smallest information set J_T . In both parts, using more information results in better combinations in terms of the expected MSE. The second formulation is in terms of the underlying optimization problems. The final formulation is in terms of the optimal solutions and is essentially a Jensen-type inequality for a scalar function of a matrix argument. Part (c) has only one formulation in terms of the underlying optimization problems because there is no explicit solution in this case.

2.3 Practical considerations and the operational strategy

In practice, of course, the conditionally optimal weights $\boldsymbol{w}^*(I_T)$ given by (4) need to be estimated, which involves estimating Σ_ξ and $\boldsymbol{b}_T$. The estimated version

$$\widehat{\boldsymbol{w}}^*(I_T) = \frac{[\widehat{\Sigma}_\xi + \widehat{\boldsymbol{b}}_T\widehat{\boldsymbol{b}}_T']^{-1}\boldsymbol{\iota}}{\boldsymbol{\iota}'[\widehat{\Sigma}_\xi + \widehat{\boldsymbol{b}}_T\widehat{\boldsymbol{b}}_T']^{-1}\boldsymbol{\iota}}, \tag{8}$$

will inherit the estimation issues from $\widehat{\Sigma}_\xi$ and $\widehat{\boldsymbol{b}}_T$. These issues may be particularly severe in the case of $\widehat{\boldsymbol{b}}_T$, which is the predictable part of $\boldsymbol{e}_{T+h}$, since the forecast model specifications should attempt to minimize this component to the greatest extent possible. Our operational strategy is to use several modifications to moderate these issues.

1. Since correlation estimation is unstable, one can employ a shrinkage technique to stabilize it, e.g. use

   $$\widetilde{\Sigma}_\xi = \alpha\Sigma_0 + (1-\alpha)\widehat{\Sigma}_\xi.$$

   The stabilizing matrix Σ_0 can be a diagonal matrix of the variances (i.e., shrinking all correlations to zero), or it can be an identity matrix $\mathbf{I}$ (i.e., shrinking the weights toward equal weights), which are the approaches often used for the unconditional weights. We denote the weight that is produced by this modification as

   $$\widehat{\boldsymbol{w}}^*_{\mathrm{COS}}(I_T) = \frac{[\widetilde{\Sigma}_\xi + \widehat{\boldsymbol{b}}_T\widehat{\boldsymbol{b}}_T']^{-1}\boldsymbol{\iota}}{\boldsymbol{\iota}'[\widetilde{\Sigma}_\xi + \widehat{\boldsymbol{b}}_T\widehat{\boldsymbol{b}}_T']^{-1}\boldsymbol{\iota}} \tag{9}$$

   and refer to $\widehat{\boldsymbol{w}}^*_{\mathrm{COS}}(I_T)$ as *conditional optimal weights with shrinkage* and explicitly specify the shrinkage parameter, e.g., $\alpha = 0.5$.

2. If the conditional bias part is more important than the remainder, then we can concentrate on using $\widehat{\boldsymbol{b}}_T$ and ignore $\widehat{\Sigma}_\xi$ and its estimation issues, i.e., use

   $$\widehat{\boldsymbol{w}}^*_b(I_T) = \frac{[\widehat{\boldsymbol{b}}_T\widehat{\boldsymbol{b}}_T']^{-1}\boldsymbol{\iota}}{\boldsymbol{\iota}'[\widehat{\boldsymbol{b}}_T\widehat{\boldsymbol{b}}_T']^{-1}\boldsymbol{\iota}}.$$

Matrix $\widehat{\boldsymbol{b}}_T\widehat{\boldsymbol{b}}_T'$ will have off-diagonal elements that can present a similar problem as the off-diagonal elements of $\widehat{\Sigma}_\xi$. If the same fix is applied and the off-diagonal elements are shrunk to zero, then we can use matrix

$$\beta B + (1-\beta)\widehat{\boldsymbol{b}}_T\widehat{\boldsymbol{b}}_T',$$

where the stabilizing matrix $B = \mathrm{diag}(\widehat{b}_{1,T}^2, \ldots, \widehat{b}_{n,T}^2)$ has squared elements of $\widehat{\boldsymbol{b}}_T = (\widehat{b}_{1,T}, \ldots, \widehat{b}_{n,T})'$ on the diagonal and zeros everywhere else. In the case when matrix $\widehat{\boldsymbol{b}}_T\widehat{\boldsymbol{b}}_T'$ is fully shrunk toward its diagonal (i.e., $\beta = 1$), the estimated conditional optimal weights will be

$$\widehat{\boldsymbol{w}}_{\mathrm{PB}}^*(I_T) = \frac{1}{\sum_{l=1}^n \widehat{b}_{l,T}^{-2}} \left(\widehat{b}_{1,T}^{-2}, \ldots, \widehat{b}_{n,T}^{-2}\right)'. \tag{10}$$

The individual weights are the inverse of the corresponding squared biases, which is very intuitive as this will place larger weights on the forecasts with small biases and smaller weights on the forecasts with large biases. We will refer to $\widehat{\boldsymbol{w}}_{\mathrm{PB}}^*(I_T)$ as *predicted bias weights*.

3. Form (10) opens up numerous possibilities to allocate the weights according to the biases of the corresponding forecasts. One option is to use the exponential function

$$\widehat{\boldsymbol{w}}_{\mathrm{PE}}^*(I_T) = \frac{1}{\sum_{l=1}^n \exp(-\gamma \widehat{b}_{l,T}^2)} \left(\exp(-\gamma \widehat{b}_{1,T}^2), \ldots, \exp(-\gamma \widehat{b}_{n,T}^2)\right)', \tag{11}$$

such that the weight decreases faster when the bias increases. The exponential function is also bounded when the bias is close to zero, which should have a positive effect on the weight performance. We will refer to $\widehat{\boldsymbol{w}}_{\mathrm{PE}}^*(I_T)$ as *predicted exponential weights* with explicitly specified parameter γ, e.g., $\gamma = 5$.

3 Real-time forecasting with conditionally optimal weights

In this section, we subject the operational strategies based on conditionally optimal weights to a real-time forecasting test using US inflation data.

3.1 Data

For our main forecasting experiments, we use data from the Philadelphia Federal Reserve's Real-Time Macroeconomic data set.[6] The measures of inflation that we consider are constructed using the Price Index for Personal Consumption Expenditure (PCE) and the gross domestic product (GDP) deflator. These measures are chosen because real-time data are available dating back to 1965Q4, which allows for the longest possible out-of-sample forecasting period. Quarterly inflation is defined as

$$\pi_t = \ln\left(\frac{p_t}{p_{t-1}}\right) \times 400,$$

where p_t is the price index.

The variables used to forecast inflation and to predict model performance are constructed from the real GDP and the civilian unemployment rate measures available in the real-time data set. The real GDP measure is used to create three predictors: 1) GDP growth, constructed as log-differenced GDP; 2) output gap, constructed using the standard Hodrick Prescott (HP) filter; 3) and a growth gap measure, which is constructed as the difference of the current GDP growth rate from the maximum growth rate observed over the previous twelve quarters. The unemployment rate is used to create two predictors: 1) the simple level and 2) as a one-sided unemployment gap measure similar to the GDP growth gap. The unemployment gap and growth gap measures follow Stock

[6] A detailed description of the data set and an explanation of its usefulness for evaluating forecasting strategies are provided by Croushore and Stark (2001).

Univariate	Phillips Curve	Direct Forecasts
AR(1)	PC Output Gap	DF Output Gap
AR(2)	PC Unemployment Gap	DF Unemployment Gap
AR(4)	PC GDP Growth	DF GDP Growth
ARMA$(1,1)$	PC Growth Gap	DF Growth Gap
ARMA$(4,4)$	PC Unemployment Rate	DF Unemployment Rate
AO	VAR All	

Table 1: Forecast model specifications.

and Watson (2010) and provide one-sided measures of the business cycle to capture the possible nonlinearity of the Phillips curve.

3.2 Models

The list of considered models is presented in Table 1. The included univariate models are chosen either because they are frequently used as benchmarks in the inflation forecasting literature or to provide variety in the specifications.[7] The AO forecast is based on the naïve random walk model employed in Atkeson and Ohanian (2001) and is the average of the previous four quarters of inflation[8]

$$\hat{\pi}^{AO}_{t+h} = \frac{1}{4}\sum_{i=1}^{4} \pi_{t-i}. \tag{12}$$

The Phillips curve (PC) specifications are bi-variate VARs with two lags of inflation and two lags of a real activity measure. VARs are used as the PC-type forecasts to provide a degree of generality to the results. The VARs do not impose a specific theory on the structure of the Phillips curve but incorporate the basic observable information that is used in many different theoretically based PC forecasts. The lag length for the VARs is selected for parsimony and is held constant throughout the exercise.

[7] For example the ARMA(1,1) is the benchmark forecast employed by Ang et al. (2007), who compared dozens of different forecast specifications covering surveys, ARMA models, regressions using real-activity measures, and term structure models; or the AR(4), which is the benchmark in Stock and Watson (2010).

[8] The AO forecast performs comparably to the inflation gap forecasting strategy proposed by Stock and Watson (2007) as shown by Faust and Wright (2013).

The specifications labeled as direct forecasts (DF) are OLS regressions of a given real activity measure on h-quarter-ahead inflation

$$\pi_{t+h} = c + \beta x_t + \epsilon_{t+h}, \tag{13}$$

where x_t represents a real activity measure and ϵ_{t+h} is the error term.[9] This specification is included primarily as a robustness check. The predictions of forecast errors are constructed using the same direct forecast specification. One concern is that the prediction step of the combination strategy is picking up a correctable form of model misspecification. The underlying assumption of the proposed combination strategy is that there is a model misspecification that can be exploited through the prediction of a model's forecast errors, but that cannot be exploited by altering the forecast model's specifications.

The VAR All model is also included for robustness. The VAR All model includes all the information that is found to predict forecast performance well into a single specification (GDP growth, output gap, and the unemployment gap). If the information used to predict a future forecast model's performance is more useful for predicting the level of inflation, then this model should provide relatively accurate forecasts. The lag length of this specification is also fixed at two for all exercises.

3.3 Forecasts and Inference

The forecast of interest in this paper is the four-quarter-ahead forecast of quarterly inflation, which is expressed as an annual rate following Ang et al. (2007). A medium horizon is studied because the predictable component of the error tends to increase with the forecast horizon and because this horizon is also one that is of particular interest to policy makers.

The real-time forecasts are constructed using the latest vintage of data available at

[9] For an explanation of the merits of direct forecasting see Marcellino et al. (2006).

each point in time. Due to lags in the release of data to the public, however, the current quarter's observation of inflation is not available at the time a forecast is made. Therefore, the considered forecast is actually the nowcast and the subsequent three quarters. The forecasts are denoted as $E_t^{\tau}\pi_{t+4}$, where t is the last observation of data and τ is the vintage of data.[10]

The forecasts are evaluated based on the root mean squared forecast error (MSFE) and the mean forecast error (MFE) to measure forecast bias. Inferences on the observed differences in MSFE are obtained using the Diebold and Mariano (1995) (DM) test for equal within-sample forecast accuracy with the Harvey et al. (1997) small sample size and long horizon correction.[11] There is not much guidance in the literature on the correct test statistic for evaluating combined forecasts. We assume that the underlying data generating process suffers from frequent structural breaks. Most test statistics are based on the assumption of asymptotic convergence to stationary distributions for the estimated regressors of the model considered and of their forecast errors.[12] Neither of which hold in this case, which is one reason for considering model combination in the first place. The use of the Diebold and Mariano test statistic follows the recommendations provided by Clark and McCracken (2013) for evaluating forecasts on real-time data and Diebold (2015), who notes that the only assumption that must be satisfied to use the DM test statistic is that the differences in squared forecast errors are covariance stationary. Inference on the bias results is obtained using a t-test with Newey and West (1987) standard errors.

The target measure of inflation to which the real-time forecasts are compared is a composite series constructed from the second release quarterly observations of inflation as they appear in the real-time data set. The use of second release data minimizes the influences of large renormalizations that occur in the sample due to definitional changes

[10] $E_t^{\tau}\pi_{t+4}$ is the f_{T+h} from Section 2.

[11] This test statistic is also occasionally referred to as the Diebold Mariano and West test statistic to reflect that it is a special case of the statistic analyzed in West (1996).

[12] See West (2006) for a review of the literature.

and provides a final measure of inflation that is closer to the actual measure that a forecaster would have been attempting to forecast at any given point in time.

3.4 Individual Model Results

Table 2 reports the real-time out-of-sample forecasting results for the individual model forecasts of PCE inflation. The results are reported for the full out-of-sample period (1970Q1-2014Q1) and for two sub-sample periods: 1983Q1-2007Q3, which roughly covers the Great Moderation, and the most recent period 2007Q4-2014Q1, which covers the Financial Crisis and the recovery. We only present the results for PCE inflation in this section because we find little difference in the forecast outcomes between the PCE and the PGDP measures of inflation. However, the PGDP results are presented in Section 4.2. All RMSFE results are reported relative to the AO forecast.

The AO forecast clearly dominates all other considered forecast specifications. The AO forecast results in the lowest RMSFE over the full sample and both subsamples with only two exceptions. The PC GDP growth forecasts and the ARMA(1,1) forecast both result in lower RMSFE during the subsample covering the Great Moderation. However, neither of the improvements are statistically different from the AO forecast. The AO forecast also has the lowest bias among the forecast models considered.

The best performing PC specification in terms of RMSFE is the specification that utilizes the one-sided growth gap measure. Although, the growth gap specification is statistically no different from the AO forecast over the full sample and the Great Moderation subsample. The worst performing PC specification in terms of RMSFE is the output gap specification, which is statistically significantly worse than the AO forecast on the full sample and on the two subsamples. Finally, all models provide unbiased forecasts for the full sample.

Individual Model Results

	1970Q1-2014Q1		1983Q1-2007Q3		2007Q4-2014Q1	
Predictor	Rel. RMSFE	Bias	Rel. RMSFE	Bias	Rel. RMSFE	Bias
Benchmark						
AO	2.353	$0.09^{\dagger}$	1.483	$0.20^{\dagger}$	2.282	$0.21^{\dagger}$
Direct Forecasts (relative RMSFE)						
DF CUR	1.287	$0.24^{\dagger}$	1.432	1.42	1.349	2.28
DF GDP	1.267	$-0.27^{\dagger}$	1.189	1.10	1.062	1.51
DF Growth Gap	1.258	$-0.08^{\dagger}$	1.262	1.27	1.109	1.67
DF Output Gap	1.368	$-0.22^{\dagger}$	1.461	1.30	1.095	1.45
DF U. Gap	1.263	$-0.09^{\dagger}$	1.246	1.22	1.153	1.79
Phillips Curves (relative RMSFE)						
PC CUR	1.064	$0.14^{\dagger}$	1.169	0.81	1.190	1.33
PC GDP growth	1.041	$-0.37^{\dagger}$	0.985	0.54	1.062	$0.65^{\dagger}$
PC Growth Gap	1.002	$-0.11^{\dagger}$	1.018	0.53	1.082	0.77
PC Output Gap	1.087	$-0.31^{\dagger}$	1.086	0.71	1.099	0.74
PC U. Gap	1.026	$-0.04^{\dagger}$	1.055	0.62	1.159	1.10
VAR ALL	1.073	$-0.31^{\dagger}$	1.198	0.54	1.090	$0.55^{\dagger}$
Univariate (relative RMSFE)						
AR(1)	1.114	$-0.22^{\dagger}$	1.066	0.84	1.078	1.05
AR(2)	1.021	$-0.15^{\dagger}$	1.017	0.64	1.073	0.85
AR(4)	1.072	$-0.30^{\dagger}$	1.019	0.63	1.067	0.82
ARMA(1, 1)	1.003	$-0.15^{\dagger}$	0.991	0.57	1.057	0.78
ARMA(4, 4)	1.085	$-0.22^{\dagger}$	1.122	0.70	1.042	0.71

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2: This table reports the individual forecast model results. The RMSFE of the AO forecast is reported in the first row. All of the remaining results are reported relative to the AO forecast. A number less than one represents an improvement in the forecast accuracy. Significance for the RMSFE results is only indicated for improvements over the benchmark. The † indicates a failure to reject the null hypothesis of unbiasedness at the 10% level.

3.5 Combined Forecasts

3.5.1 Predicting Forecast Errors

The prediction of inflation forecast errors in our application is similar to the concept of inflation gap forecasting explored by Stock and Watson (2007, 2010) and Cogley et al. (2010). The inflation gap is defined as the deviation of inflation from a stochastic trend. An inflation gap forecast is constructed by forecasting the inflation gap and then adding it back to the last observation of the trend. Stock and Watson (2010) and Faust and Wright (2013) both find that inflation gap forecasting offers modest improvements over other parsimonious time-series models in pseudo (single vintage of data) and real-time (multiple vintages of data) out-of-sample experiments, respectively. The prediction of the inflation gap is a forecast of the trend forecast's error. The difference in our case is that we use the gap prediction to aid in selecting forecast weights rather than using it to engage in intercept correction.

The prediction of a forecast model's error is obtained via a direct forecast of the model's real-time forecast error. The direct forecast regresses the four-quarter-ahead forecast error, $e_{i,t+4} = \pi_{t+4} - E_{i,t}\pi_{t+4}$, on a real activity measure

$$e_{i,t+4} = c + \beta_i x_t + \xi_{i,t+4} \tag{14}$$

for the i^{th} considered model.[13] The forecast error series is constructed using real-time errors obtained from comparing past real-time forecasts to the composite series of second release information. However, the last forecast error in each period is compared to the first release information because the second release is not available at the time the forecast

[13] The real activity measure acts as a "tester" using the terminology of Timmermann and Zhu (2016) and provides information on the relative forecast performance of the competing models. Timmermann and Zhu (2016) provide a thorough analysis of the usefulness of this type of specifications for predicting relative forecast accuracy. They, however, do not study its usefulness for combining forecasts. This specification is also similar to the specification considered by Stock and Watson (2010) to forecast the inflation gap.

is made. By construction, this procedure introduces new information into the forecasts because the individual forecast models are estimated on the most recent vintage of data available at the time the forecast is made, whereas the series of forecast errors contains information from multiple vintages. Therefore, the forecast error series incorporates some information about revisions into the final combined forecast.[14]

3.5.2 Combining forecasts with conditionally optimal weights

The out-of-sample forecasting exercise requires the data to be separated into three subsets. The required divisions are 1) a training subset to estimate the initial parameters of the forecast models, 2) an in-sample forecasting period to recursively forecast to construct an initial series of forecast errors to estimate Equation (14), and 3) an out-of-sample period to conduct out-of-sample forecasts. For the forecast experiments presented in this paper, the three periods are 1947Q2-1965Q4, 1966Q1-1969Q4, and 1970Q1-2014Q1, respectively.

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

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 (14) is estimated on the available real-time forecast errors for each of the candidate forecasts.
4. Equation (14) 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.

[14] We tested both revised and unrevised forecast error series, and the revised series do appear to add a small amount of forecasting power relative to the unrevised series.

3.5.3 Combined Forecast Results

Table 3 reports the results for combinations of all seventeen individual forecasts using the three different operational variants of conditionally optimal weights. The first combination strategy shown is conditionally optimal weights with shrinkage, Equation (9), where $\hat{\Sigma}_\xi$ is shrunk toward an identity matrix. The shrinkage parameter is set to $\alpha = 0.5$ to reduce the noise coming from the estimation uncertainty of this term. The second operational combination strategy, predicted bias weights, Equation (10), uses only the predictions of the squared bias $(\widehat{b}^2_{1,T}, \dots, \widehat{b}^2_{n,T})$ by removing $\hat{\Sigma}_\xi$ and ignores off-diagonal terms in $\widehat{\boldsymbol{b}}_T \widehat{\boldsymbol{b}}'_T$. The final operational strategy that we consider is predicted exponential weights, Equation (11), with $\gamma = 5$. The γ parameter in this specification is chosen by searching over whole number values to minimize the MSFE in a pre-sample 1967Q1-1969Q4. The benchmark for the combined forecasts is the equal weights forecast. All RMSFEs are normalized against its value. In addition, we also provide the AO forecast results in Table 3, and the results relative to the AO benchmark are presented in Table A1.

Conditionally Optimal Weights with Shrinkage significantly outperform equal weights in four of the five full sample experiments and qualitatively outperform equal weights in approximately two thirds of the forecasting experiments, yielding lower RMSFEs and reductions in forecast bias. This case also yields significant improvements relative to the benchmark AO forecast for the full sample for the output gap, unemployment gap, and unemployment rate specifications. The combination strategy, however, does not yield any improvements in the recent subsamples relative to AO (results relative to AO are shown in Table A1 in Appendix A).

The second operational case, Predicted Bias Weights, performs almost as well as conditionally optimal weights with shrinkage in terms of RMSFE on the full sample and shows improvements in the Great Moderation subsample. This case yields statically significant improvements relative to equal weights in two instances in the Great Moderation

Combined Forecast Results

	1970Q1-2014Q1		1983Q1-2007Q3		2007Q4-2014Q1	
Predictor	Rel. RMSFE	Bias	Rel. RMSFE	Bias	Rel. RMSFE	Bias
Cond. Optimal Weights with Shrinkage: $\alpha = 0.5$						
Output Gap	0.936*	$-0.35^{\dagger}$	0.960	$0.04^{\dagger}$	1.031	$0.04^{\dagger}$
U. Gap	0.850***	$-0.11^{\dagger}$	1.007	$0.19^{\dagger}$	1.028	$0.11^{\dagger}$
GDP Growth	0.899**	$-0.17^{\dagger}$	0.956	$0.07^{\dagger}$	0.991	$0.18^{\dagger}$
Growth Gap	1.004	$-0.03^{\dagger}$	0.983	$0.19^{\dagger}$	1.004	$0.27^{\dagger}$
CUR	0.878**	$0.04^{\dagger}$	0.988	$0.34^{\dagger}$	1.022	$0.09^{\dagger}$
Predicted Bias Weights						
Output Gap	0.976*	$-0.13^{\dagger}$	0.988	0.63	0.984	0.82
U. Gap	0.981	$0.08^{\dagger}$	1.014	0.79	0.999	1.00
GDP Growth	0.920***	$-0.03^{\dagger}$	0.958*	0.52	1.059	1.07
Growth Gap	0.932***	$-0.05^{\dagger}$	0.933**	0.41	1.064	1.44
CUR	0.929***	$0.09^{\dagger}$	1.002	0.64	1.050	0.96
Predicted Exponential Weights: $\gamma = 5$						
Output Gap	0.875***	$-0.19^{\dagger}$	0.899***	$0.23^{\dagger}$	0.954**	0.74
U. Gap	0.929*	$0.22^{\dagger}$	0.974	0.53	0.998	1.06
GDP Growth	0.883***	$0.15^{\dagger}$	0.922**	0.43	1.011	1.11
Growth Gap	0.958	$0.26^{\dagger}$	0.954	0.46	1.028	1.21
CUR	0.952	$0.29^{\dagger}$	0.974	0.48	1.001	1.06
Comparisons						
AO	0.949	$0.09^{\dagger}$	0.947	$0.20^{\dagger}$	0.956	$0.21^{\dagger}$
Equal weights	1.000	$-0.14^{\dagger}$	1.000	0.80	1.000	1.07

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3: The RMSFEs are shown relative to the equal weights forecast. Significance for the RMSFE results is only indicated for improvements over the benchmark. The † indicates a failure to reject the null hypothesis of unbiasedness at the 10% level.

Best Predicted Model in Each Period

	1970Q1-2014Q1		1983Q1-2007Q3		2007Q4-2014Q1	
Predictor	Rel. RMSFE	Bias	Rel. RMSFE	Bias	Rel. RMSFE	Bias
Predicted Exponential Weights: $\gamma \to \infty$						
Output Gap	0.893***	0.28†	0.949	0.22†	0.943**	0.52†
U. Gap	0.959	0.33†	1.058	0.57	1.075	1.28
GDP Growth	0.913**	0.09†	0.921**	0.26†	1.115	1.04
Growth Gap	0.976	0.14†	0.952	0.25†	1.166	1.68
CUR	0.967	0.26†	1.014	0.41†	1.044	1.09
Comparisons						
AO	0.949	0.09†	0.947	0.20†	0.956	0.21†
Equal weights	1.000	−0.14†	1.000	0.80	1.000	1.07

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4: The RMSFEs are shown relative to the equal weights forecast. Significance for the RMSFE results is only indicated for improvements over the benchmark. The † indicates a failure to reject the null hypothesis of unbiasedness at the 10% level.

subsample for predictions using GDP growth and the GDP growth gap. Compared to the AO forecast, its absolute performance is similar to the previous case with slightly weaker performance in the full sample but better performance in the subsamples.

The third operational strategy, Predicted Exponential Weights, is our preferred strategy. This strategy uses the same information as the previous strategy but incorporates shrinkage in a nonstandard way. The weights in this case are not shrunk toward equal weights but rather increased toward the forecast with the lowest expected squared error. This strategy performs the best. It yields consistent improvement relative to equal weights in the full sample and across both subsamples with statically significant improvements for at least one specification in each sample considered. Compared to the AO forecast, the results are good with qualitative improvement in all samples considered.[15]

The economic significance of the improvements in the Predicted Exponential Weight

[15]The best overall predictor for the weights observed in these exercises is the output gap measure. This is somewhat surprising because Orphanides and van Norden (2005) show that the HP-filtered output gap has very little predictive power over inflation in real time and because the output gap is the worst predictor of inflation on average among the PC and DF forecasts reported in Table 2. One explanation for this finding is that the HP filter provides an estimate of the output gap that only captures large business cycle fluctuation, which is when a change in relative efficiency between the Phillips curve and univariate models is most likely.

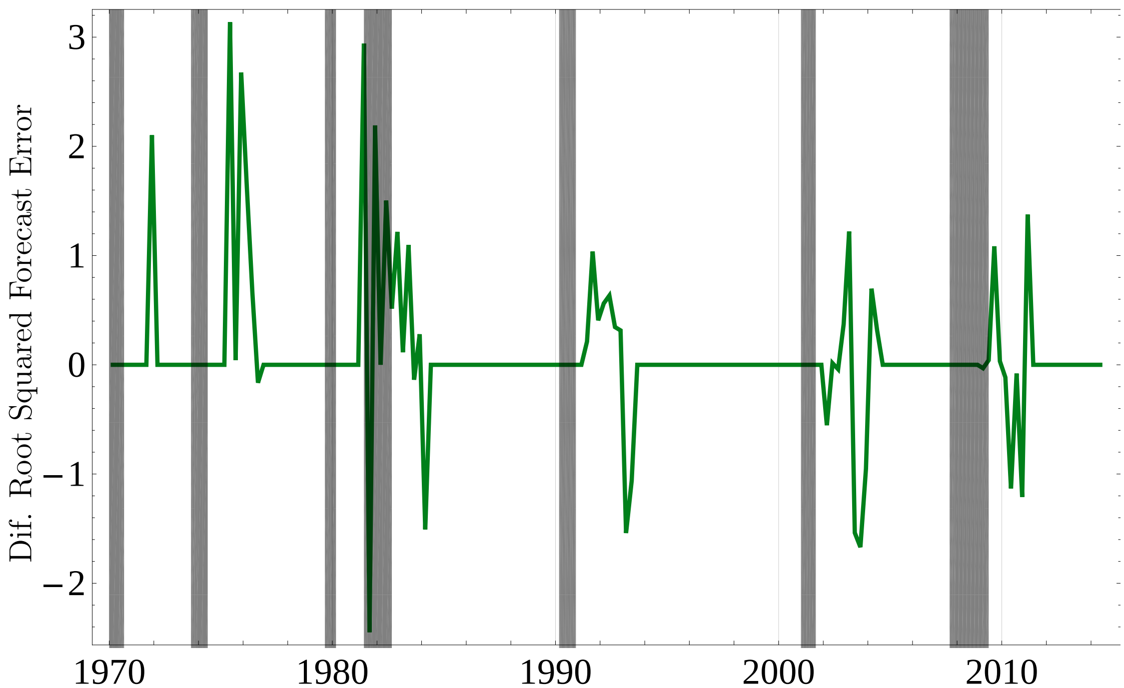

Figure 1: This figure plots the difference in root squared forecast error between the AO and the Exponential Weights forecast using the output gap as a predictor (positive values indicate that Exponential Weights forecast is more accurate). The shaded bars represent the NBER recession dates.

case is illustrated in Figure 1. This figure plots the difference in the root squared forecast error between the AO forecast and the Predicted Exponential Weight forecast that uses the output gap as the predictor for each period. Points that are above zero correspond to better forecast accuracy. Due to the choice of the shrinkage parameter γ, the difference between the forecasts of the two combination strategies is small in most periods because almost all weight is placed on the AO forecast. However, following recessions the combined forecast yields improvements over the AO forecast that average approximately one percentage point of inflation. The timing of these improvements is significant. Stock and Watson (2009) show that the forecast accuracy of the Phillips curve relationship is episodic. It tends to increase in accuracy during downturns and decrease in times of economic strength. The forward-looking weights used here are predicting these changes and shifting the weights accordingly in real time.

Table 4 shows the limiting case of Predicted Exponential Weights. The parameter γ

is sent to infinity, which is equivalent to placing a weight of one on the best predicted forecast in each time period removing the hedging advantage of a combined forecast. This case is of particular interest because, as noted in Timmermann (2006), choosing a single model in every period typically results in very poor out-of-sample forecasting accuracy. Therefore, the results here are quite surprising. Choosing the expected best model in each period actually leads to reductions in the relative RMSFE compared to equal weights in a majority of the out-of-sample forecast experiments and even results in a lower RMSFE compared to the AO benchmark for some predictors in every sample period considered (see Table A2 in the Appendix A). It also leads to a reduction in bias compared to the combined forecasts in most cases. The results illustrate that the gains in forecast efficiency observed in Table 3 are driven by correct predictions of the actual best performing forecast model in each period.

4 Further Analysis and Robustness

4.1 Forward versus Backward-Looking Weights

As noted in the previous section, 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 (the models in the first two columns of Table 1). We use the Predicted Exponential Weight given by Equation (11) and a backward-looking modification

$$\widehat{\boldsymbol{w}}^{*}_{\text{AO}}(I_T) = \frac{1}{\sum_{l=1}^{n} \exp(-\gamma \widetilde{b}^2_{l,T})} \Big(\exp(-\gamma \widetilde{b}^2_{1,T}), \ldots, \exp(-\gamma \widetilde{b}^2_{n,T})\Big)', \tag{15}$$

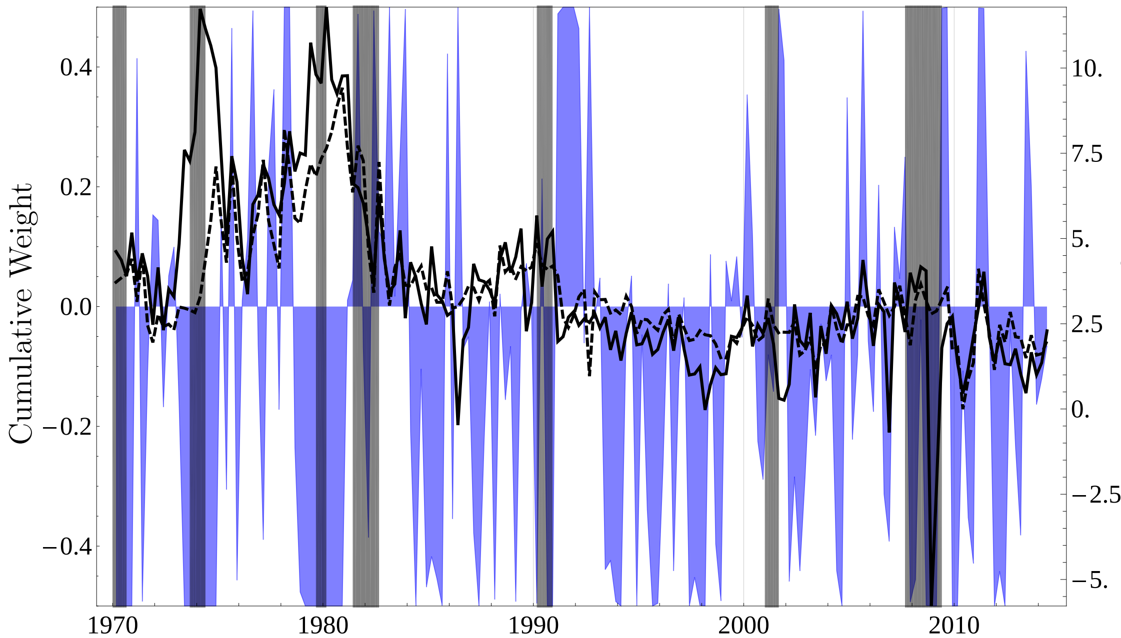

Figure 2: This figure depicts the PCE measure of US inflation from 1970Q1 to 2014Q1 (solid), the ex post predicted weights forecast (dashed), and the cumulative weight placed on the Phillips curve forecasts relative to equal weights for the ex post predicted weights (shaded blue). The dark bars indicate the NBER recession dates.

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. Figure 2 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.[16] The cumulative weights illustrated

[16]The AO and equal weights combined forecasts have a relative RMSFE of 1.0004. The benchmark

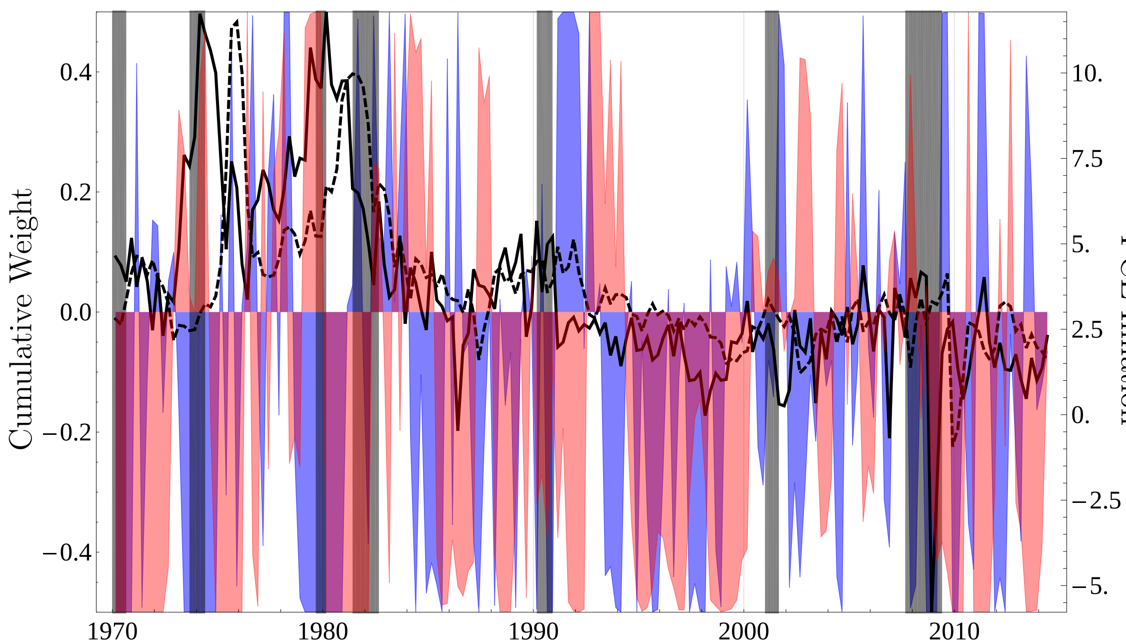

Figure 3: This figure depicts the PCE measure of US inflation from 1970Q1 to 2014Q1 (solid), the AO weights combined forecast (dashed), the cumulative ex post weights relative to equal weights (shaded blue), and the cumulative AO weights relative to equal weights (shaded red). The dark bars indicate the NBER recession dates.

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 3 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 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 where the PC

combined weights also result in a slight increase in RMSFE compared to actually forecasting with the ex post best model in each period.

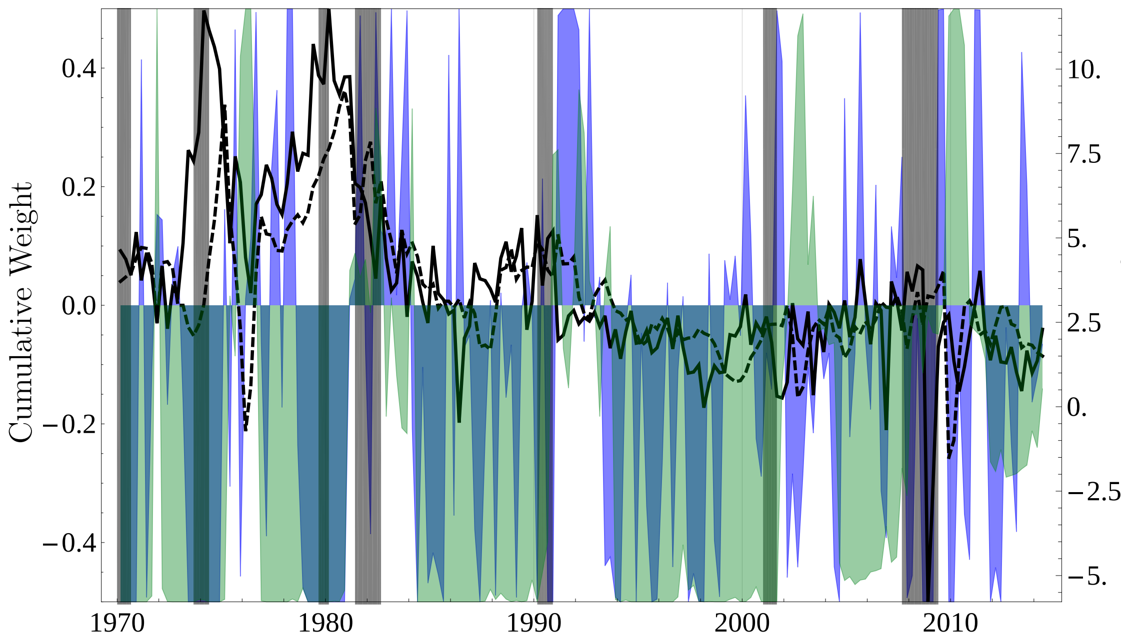

Figure 4: This figure shows the PCE measure of US inflation from 1970Q1 to 2014Q1 (solid), the Predicted Exponential Weights combined forecast (dashed), the cumulative ex post weights relative to equal weights (shaded blue), and the cumulative Predicted Exponential Weights relative to equal weights (shaded green). The dark bars indicate the NBER recession dates.

forecasts perform well, which is of course precisely when the strategy is about to lose forecast efficiency relative to the univariate forecasts.

Figures 2 and 3 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 4 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 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.

4.2 Forecasting Tournament

Comparisons of combined forecast techniques face an external validity problem because the results are sensitive to the set of forecast models considered. This concern is particularly relevant when comparing a combination strategy to an equal weights forecast. The equal weights strategy clearly has no mechanism to filter out obviously poor forecasts. Therefore, it is easy to construct a straw man equal weights forecast by considering poor performing forecasts that a sophisticated model combination strategy can easily detect and which actual forecasters would not consider. To overcome this issue and provide a relatively fair comparison of equal weights to the proposed predicted weights strategy, we conduct a forecast tournament that varies the set of models in the combined forecasts. The tournament is conducted by selecting the twelve most efficient models from the 2007Q4-2014Q1 subsample found in Table 2 and then considering every combination of the twelve distinct forecasts taken n at time, where $n = 2, 3, ..., 12$. This provides 4,083 different sets of models to combine.[17]

For the tournament, we compare four different forecast combination strategies:[18]

1. Equal weights
2. Weights based on the past observed MSFE of each model following Stock and Watson (2004), where

$$\widehat{\boldsymbol{w}}_{\text{SW}} = \frac{1}{\sum_{i=1}^{n} MSFE_{i,t}^{-k}} \left(MSFE_{1,t}^{-k}, ..., MSFE_{n,t}^{-k}\right)' \tag{16}$$

[17]The number of distinct combinations of the twelve models for each n is as follows: $n = 2 \rightarrow 66$ sets, $n = 3 \rightarrow 220$ sets, $n = 4 \rightarrow 495$ sets, $n = 5 \rightarrow 792$ sets, $n = 6 \rightarrow 924$ sets, $n = 7 \rightarrow 792$ sets, $n = 8 \rightarrow 495$ sets, $n = 9 \rightarrow 220$ sets, $n = 10 \rightarrow 66$ sets, $n = 11 \rightarrow 12$ sets, $n = 12 \rightarrow 1$ set.

[18]We also considered the optimal weights implied by regressing all past forecasts on the actual realization of inflation proposed by Granger and Ramanathan (1984). However, the weights perform so poorly compared to the other four methods considered that it did not provide a useful comparison.

such that $MSFE_{i,t} = (1/m)\sum_{\tau=t-m}^{t} e_{i,\tau-4}^2$, m is the sample size, and $k = 5$ is a shrinkage parameter. The MSFE weights are one of the simple combination procedures that, similar to equal weights, consistently improve upon more sophisticated weighting procedures.

3. The AO weights discussed in the previous section, Equation (15), with $\gamma = 5$.

4. Predicted Exponential Weights using the output gap to predict errors, Equation (11), with $\gamma = 5$.

Figure 5 presents a summary of the results for the real-time out-of-sample forecasting exercises conducted on the 1970Q1-2014Q1 PCE and PGDP measures of inflation. This figure shows the median, minimum, and maximum RMSFE observed for each subset of models of size n relative to the AO forecast RMSFE. The maximum RMSFE plot shows the worst case scenario for each of the combination methods for combining n different forecasts, the minimum RMSFE plot shows the best case scenario for combining n different forecasts, and the median RMSFE provides a measure of the distribution of RMSFE observed for combining n different forecasts.

The maximum RMSFE results show that each combination method exhibits approximately equal risk over the full sample. The maximum RMSFE or worst forecasting outcomes of all four strategies are comparable to each other across all sets of size n. Accuracy is increasing in the number of forecasts considered in all cases.

The minimum RMSFE results show a clear advantage for the predicted weights strategy. The predicted weights strategy consistently results in the lowest observed RMSFE among the four different forecast combination strategies. The PGDP results are particularly impressive with consistent improvements approaching 15% relative to the AO forecast for the minimum RMSFE observations for all n.

The median RMSFE results presented in Figure 5 also show an advantage for the predicted weights forecasts. The median improvements in efficiency are as high as 10%

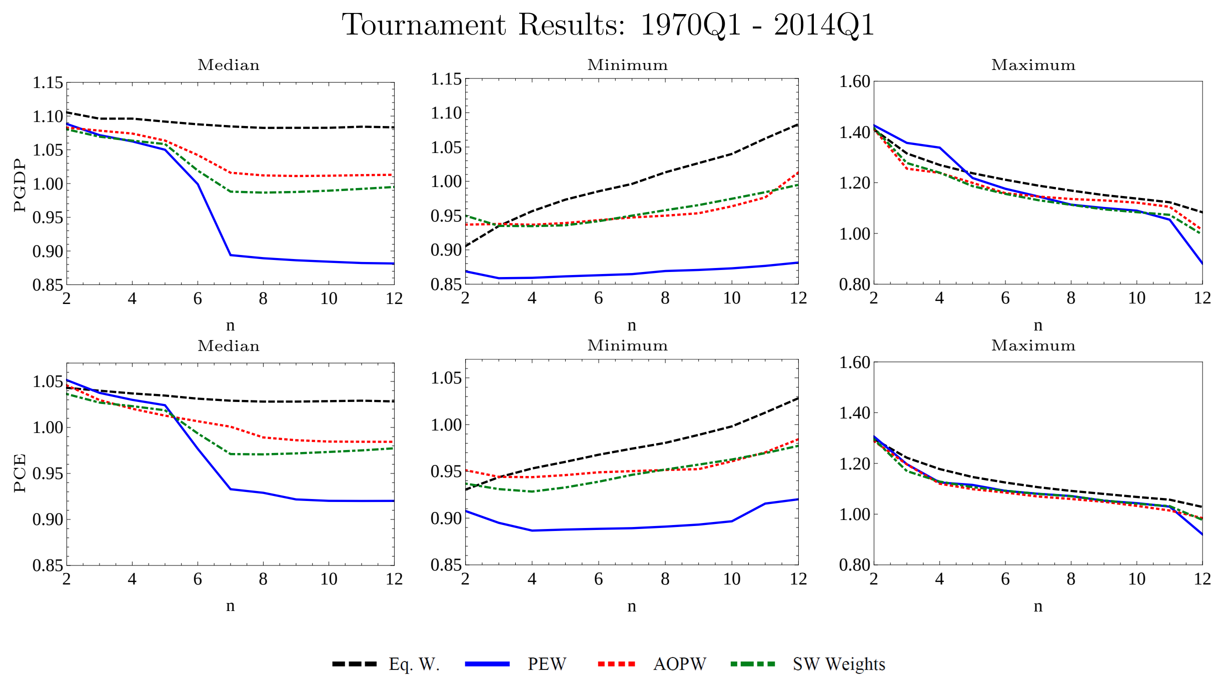

Figure 5: Median, minimum, and maximum relative RMSFE results for combinations of n different models. The results are relative to the RMSFE of the AO forecast.

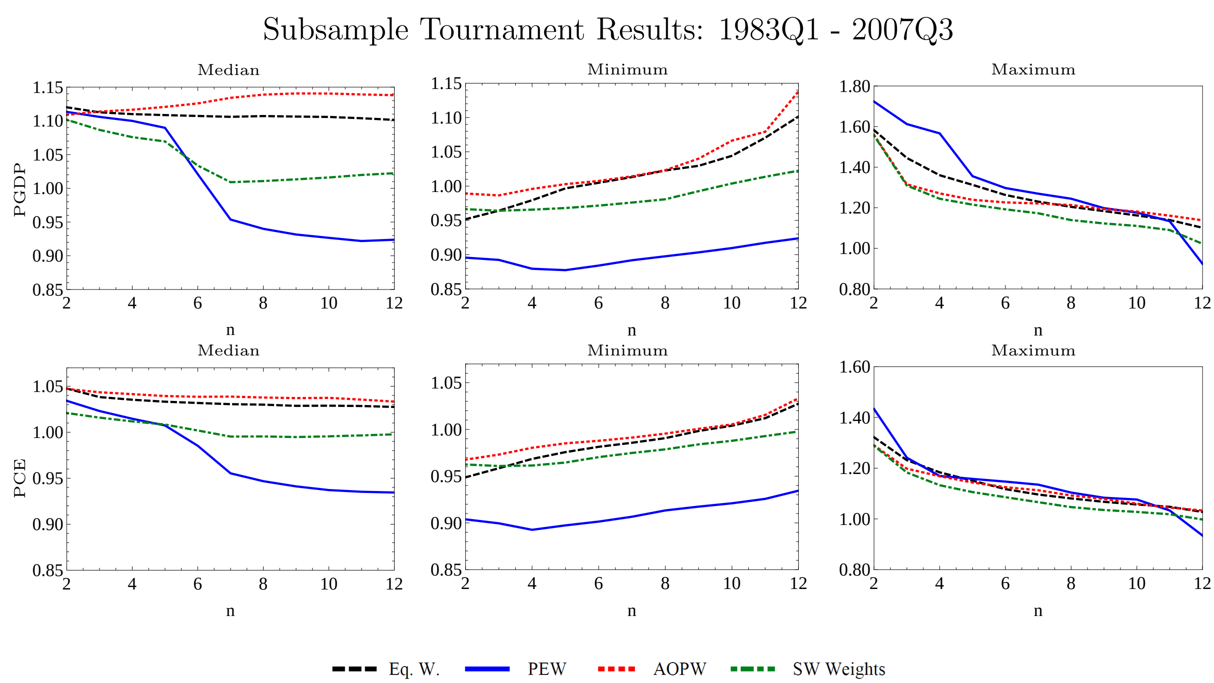

Figure 6: Median, minimum, and maximum relative RMSFE results for combinations of n different models. The results are relative to the RMSFE of the AO forecast.

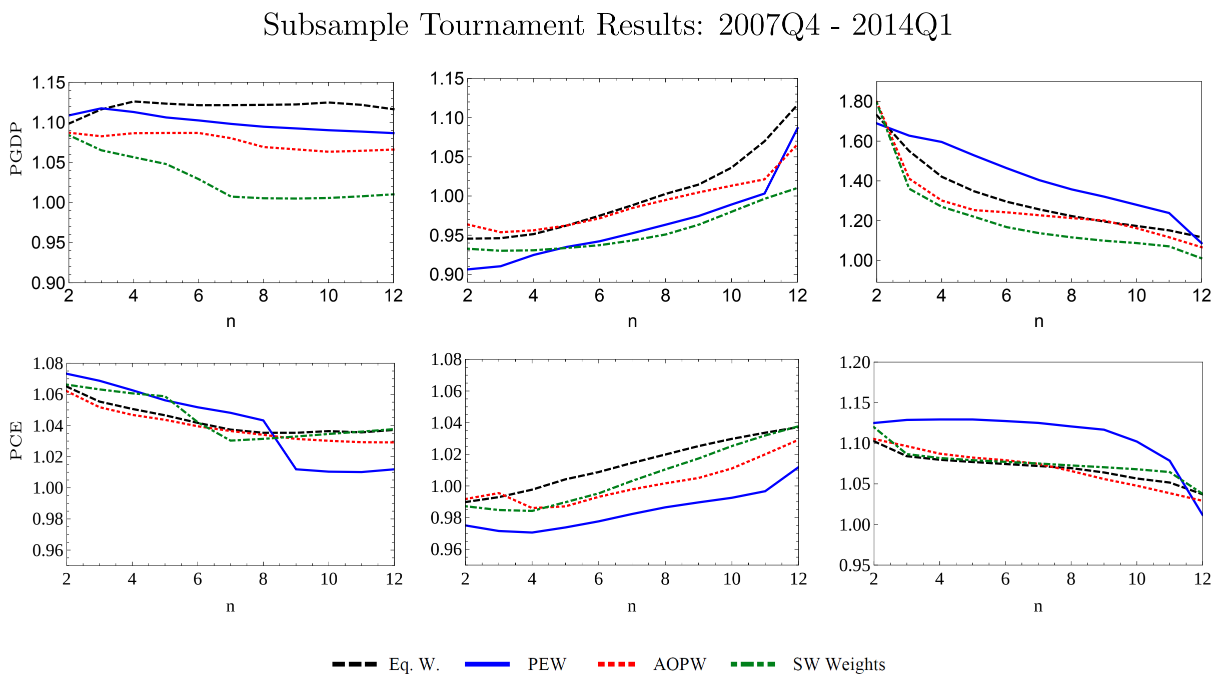

Figure 7: Median, minimum, and maximum relative RMSFE results for combinations of n different models. The results are relative to the RMSFE of the AO forecast.

relative to the AO forecast for combinations of $n > 5$ models. The increase in efficiency at $n > 5$ also provides some evidence of the exploitable time-varying trade-off between PC and univariate forecasts. For $n > 5$, almost all experiments include at least one PC and one univariate forecast model.

Figures 6 and 7 present the results for the two subsamples. The 1983Q1-2007Q4 subsample results are consistent with the full sample results. The 2007Q4-2014Q1 subsample results, however, are attenuated compared to the RMSFEs obtained on the other samples. The attenuation is most pronounced for the PGDP measure of inflation. Here, there are no consistent improvements over equal weights, although subsequent exploration revealed that increasing the γ parameter can improve performance in this sample (shown in Figure A1 in Appendix A).

The results shown in Figures 5, 6, and 7 can be approximately replicated using either the GDP growth or the unemployment gap predictors to construct the predicted weights. The GDP growth measure in particular produces forecasts that are comparable to the output gap forecasts across all considered samples. The results for the one-sided growth gap and unemployment rate measure, however, are less impressive. These predictors perform well compared to the equal weights combined forecasts but often fail to outperform the AO weights and SW weights.

4.3 Intercept Correction

Here, we examine the real-time out-of-sample forecasting performance of the individual models if the predicted errors are used for intercept correction. This exercise allows us to assess whether the information that we exploit in the weights should actually just be used to improve the individual forecast specifications rather than to construct weights. Table A3 presents the intercept correction results. The intercept correction results use the output gap predictions of the forecast errors to correct the point forecasts of each model such that

$$E_t \pi_{i,t+4}^{IC} = E_t \pi_{i,t+4} + E_t e_{i,t+4}. \tag{17}$$

Table A3 shows that the point forecasts of the forecast errors are not very accurate. The intercept corrected forecasts are less efficient than the uncorrected forecasts in almost all cases. The one exception is the AO forecast, although, the reported improvement in forecast efficiency is not statistically different from the uncorrected AO forecast. The results are similar for all of the five real activity measures considered to predict forecast errors.

The explanation for the disparity in the effectiveness between predicted weights forecasts and the intercept corrected forecasts is that the two strategies use the predictions of the forecast errors in different ways. Predicted weights exploits the relative ranking of the forecasts implied by the predicted forecast errors, whereas intercept correction relies on the accuracy of the actual point forecast of the error. The differences suggest that the predictions of forecast errors contain information about relative performance but little information about the absolute performance of the considered models.

4.4 New Zealand Inflation

Here, we reproduce all of our results using real-time data from New Zealand. The US and New Zealand economies clearly differ along a number of key dimensions, but inflation is difficult to predict in both countries. Therefore, the New Zealand data act as a check on whether we are exploiting correlations that are specific to US data. The real-time New Zealand data come from a dataset provided by the Reserve Bank of New Zealand. The dataset does not include real-time measures of PCE or PGDP; thus, quarterly non-tradable CPI inflation is used instead.[19] Inflation is defined and computed using the same definitions employed for the US data and the same real GDP and unemployment measures are constructed with one exception. The New Zealand dataset contains a real-time estimate of the output gap. We use this in place of the GDP growth gap.

Table A4 reports the results for Predicted Exponential Weights for the forecasts in Table 1 used in Section 3.5. The out-of-sample forecast period for New Zealand is 1997Q1-2014Q1. The table shows that the results are similar to those obtained on US data. Predicted Exponential Weights results in significant increases in forecast efficiency relative to the AO forecast and performs about as well as the equal weights forecasts for combinations of all 17 models.

Figure 8 presents the results for the forecast tournament on New Zealand real-time data. The results here are similar to the US results. Predicted Exponential Weights outperforms the three other strategies, robustly producing a lower RMSFE on average.

[19]Since New Zealand is a small open economy, the measure of inflation that has the most similar relationship with the output gap as PCE and PGDP for US data is non-tradable CPI inflation.

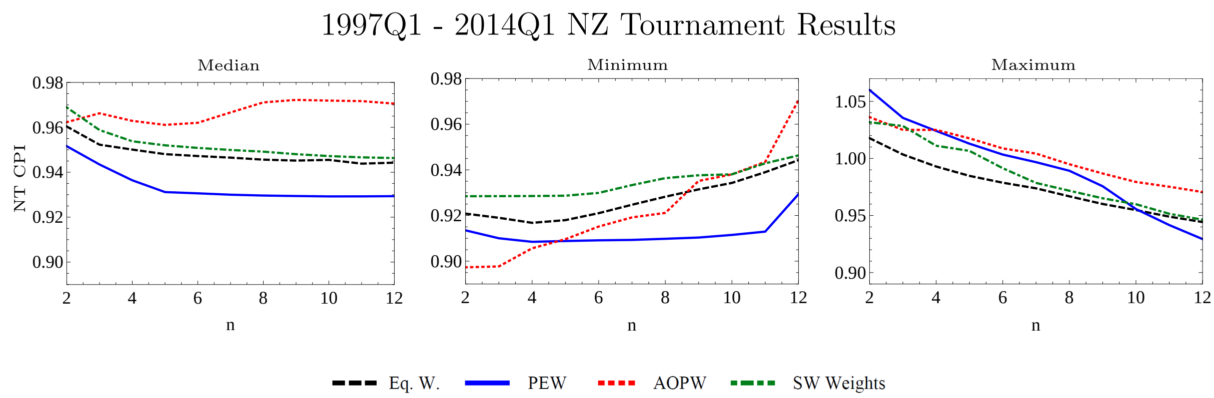

Figure 8: Median, minimum, and maximum relative RMSFE results for combinations of n different models with $\gamma = 5$ for New Zealand non-tradable inflation. The results are relative to the RMSFE of the AO forecast.

5 Conclusion

We showed that when there is predictable information in forecast errors, a combined forecast should be constructed to minimize a conditional expected loss function. We proved that forecast combinations constructed in this way improve upon unconditional combinations commonly used in the literature and that the improvements are increasing when more information becomes available. Our theoretical findings support forward-looking approaches to combining forecasts, where forecasts are weighted by their expected performance rather than their past performance.

We evaluate our forward-looking approaches using a real-time forecasting experiment of US and New Zealand inflation and find that they perform well against equally weighted forecasts and the common benchmark of a random walk. In particular, the Predicted Exponential Weights strategy provides robust improvements over equal weights in a variety of settings including different subsamples and when varying the forecast pool. Our empirical results represent a proof-of-concept that forward-looking approaches to combining forecasts are less susceptible to the issues surrounding the forecast combination puzzle than backward-looking strategies. In forecasting, one should evaluate conditional expected performance and use forward-looking approaches to combine forecasts.

References

Aiolfi, M. and A. Timmermann (2006). Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics 135*(1-2), 31–53.

Ang, A., G. Bekaert, and M. Wei (2007). Do macro variables, asset markets, or surveys forecast inflation better? *Journal of Monetary Economics 54*(4), 1163–1212.

Atkeson, A. and L. E. Ohanian (2001). Are phillips curves useful for forecasting inflation? *Federal Reserve bank of Minneapolis quarterly review 25*(1), 2–11.

Bates, J. M. and C. W. J. Granger (1969). The combination of forecasts. *Operations Research Quarterly 20*(4), 451–468.

Claeskens, G., J. R. Magnus, A. L. Vasnev, and W. Wang (2016). The forecast combination puzzle: A simple theoretical explanation. *International Journal of Forecasting 32*(3), 754–762.

Clark, T. and M. McCracken (2013). Advances in forecast evaluation. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting, Volume 2B*, pp. 1107–1201. Elsevier, Horth-Holland.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting 5*(4), 559–583.

Clements, M. P. and D. F. Hendry (1996). Intercept corrections and structural change. *Journal of Applied Econometrics 11*(5), 475–494.

Cogley, T., G. E. Primiceri, and T. J. Sargent (2010). Inflation-gap persistence in the US. *American Economic Journal: Macroeconomics 2*(1), 43–69.

Croushore, D. and T. Stark (2001). A real-time data set for macroeconomists. *Journal of Econometrics 105*(1), 111–130.

Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold-Mariano tests. *Journal of Business & Economic Statistics 33*(1), 1–9.

Diebold, F. X. and J. A. Lopez (1996). Forecast evaluation and combination. *Handbook of Statistics 14*, 241–268.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 134–144.

Elliott, G. (2011). Averaging and the optimal combination of forecasts. *UCSD working paper*, available at econweb.ucsd.edu/∼grelliott/AveragingOptimal.pdf.

Elliott, G. and A. Timmermann (2016). *Economic Forecasting.* Princeton University Press.

Faust, J. and J. H. Wright (2013). Forecasting inflation. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting, Volume 2A*, pp. 2–56. Elsevier, Horth-Holland.

Giacomini, R. and B. Rossi (2009). Detecting and predicting forecast breakdowns. *The Review of Economic Studies 76*(2), 669–705.

Granger, C. W. J. (1989). Invited review combining forecasts—twenty years later. *Journal of Forecasting 8*(3), 167–173.

Granger, C. W. J. and R. Ramanathan (1984). Improved methods of combining forecasts. *Journal of Forecasting 3*(2), 197–204.

Harvey, D., S. Leybourne, and P. Newbold (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting 13*(2), 281–291.

Hendry, D. F. and M. P. Clements (2004). Pooling of forecasts. *The Econometrics Journal 7*(1), 1–31.

Marcellino, M., J. H. Stock, and M. W. Watson (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics 135*(1-2), 499–526.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Orphanides, A. and S. van Norden (2005). The reliability of inflation forecasts based on output gap estimates in real time. *Journal of Money, Credit and Banking 37*(3), 583–601.

Smith, J. and K. F. Wallis (2009). A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics 71*(3), 331–355.

Stock, J. H. and M. W. Watson (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature 41*(3), 788–829.

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.

Stock, J. H. and M. W. Watson (2007). Why has US inflation become harder to forecast? *Journal of Money, Credit and Banking 39*, 3–33.

Stock, J. H. and M. W. Watson (2009). Phillips curve inflation forecasts. In J. Fuhrer, Y. K. Kodrzycki, J. S. Little, and G. P. Olivei (Eds.), *Understanding Inflation and the Implications for Monetary Policy*, pp. 101–186. Cambridge: MIT Press.

Stock, J. H. and M. W. Watson (2010). Modeling inflation after the crisis. *NBER Working Paper Series*, Working Paper 16488.

Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, pp. 135–196. Elsevier, Horth-Holland.

Timmermann, A. and Y. Zhu (2016). Monitoring forecasting performance. *UCSD working paper*.

Wallis, K. F. (2011). Combining forecasts - forty years later. *Applied Financial Economics 21*(1-2), 33–41.

Wallis, K. F. and J. D. Whitley (1991). Sources of error in forecasts and expectations: UK economic models, 1984-88. *Journal of Forecasting 10*(3), 231–253.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica 64*(5), 1067–1084.

West, K. D. (2006). Forecast evaluation. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, pp. 99–134. Elsevier, Horth-Holland.

Appendix A

Combined Forecast Results

	1970Q1-2014Q1		1983Q1-2007Q3		2007Q4-2014Q1	
Predictor	Rel. RMSFE	Bias	Rel. RMSFE	Bias	Rel. RMSFE	Bias
Cond. Optimal Weights with Shrinkage: $\alpha = 0.5$						
Output Gap	0.986	$-0.35^{\dagger}$	1.014	$0.04^{\dagger}$	1.079	$0.04^{\dagger}$
U. Gap	0.985***	$-0.11^{\dagger}$	1.063	$0.19^{\dagger}$	1.075	$0.11^{\dagger}$
GDP Growth	0.947*	$-0.17^{\dagger}$	1.009	$0.07^{\dagger}$	1.037	$0.18^{\dagger}$
Growth Gap	1.058	$-0.03^{\dagger}$	1.038	$0.19^{\dagger}$	1.050	$0.27^{\dagger}$
CUR	0.926*	$0.04^{\dagger}$	1.044	$0.34^{\dagger}$	1.070	$0.09^{\dagger}$
Predicted Bias Weights						
Output Gap	1.028	$-0.13^{\dagger}$	1.044	0.63	1.030	0.82
U. Gap	1.034	$0.08^{\dagger}$	1.070	0.79	1.046	1.00
GDP Growth	0.970	$-0.03^{\dagger}$	1.011	0.52	1.108	1.07
Growth Gap	0.982	$-0.05^{\dagger}$	0.986	0.41	1.113	1.44
CUR	0.979	$0.09^{\dagger}$	1.058	0.64	1.099	0.96
Predicted Exponential Weights: $\gamma = 5$						
Output Gap	0.922***	$-0.19^{\dagger}$	0.949	$0.23^{\dagger}$	0.998	0.74
U. Gap	0.979	$0.22^{\dagger}$	1.028	0.53	1.044	1.06
GDP Growth	0.931***	$0.15^{\dagger}$	0.974	0.43	1.058	1.11
Growth Gap	1.009	$0.26^{\dagger}$	1.007	0.46	1.076	1.21
CUR	1.003	$0.29^{\dagger}$	1.028	0.48	1.047	1.06
Equal Weights						
N/A	1.054	$-0.14^{\dagger}$	1.056	0.80	1.046	1.07
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$						

Table A1: The RMSFEs are shown relative to the AO forecast. Significance for the RMSFE results is only indicated for improvements over the benchmark. The † indicates a failure to reject the null hypothesis of unbiasedness at the 10% level.

Best Predicted Model in Each Period						
	1970Q1-2014Q1		1983Q1-2007Q3		2007Q4-2014Q1	
Predictor	Rel. RMSFE	Bias	Rel. RMSFE	Bias	Rel. RMSFE	Bias
Predicted Exponential Weights: $\gamma \to \infty$						
Output Gap	0.941**	0.28†	1.002	0.22†	0.987	0.52†
U. Gap	1.011	0.33†	1.117	0.57	1.124	1.28
GDP Growth	0.962	0.09†	0.973	0.26†	1.167	1.04
Growth Gap	1.029	0.14†	1.005	0.25†	1.220	1.68
CUR	1.019	0.26†	1.070	0.41†	1.093	1.09
Equal Weights						
N/A	1.054	−0.14†	1.056	0.80	1.046	1.07
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$						

Table A2: The RMSFEs are shown relative to the AO forecast. Significance for the RMSFE results is only indicated for improvements over the benchmark. The † indicates a failure to reject the null hypothesis of unbiasedness at the 10% level.

2007Q4 - 2014Q1 Tournament Results

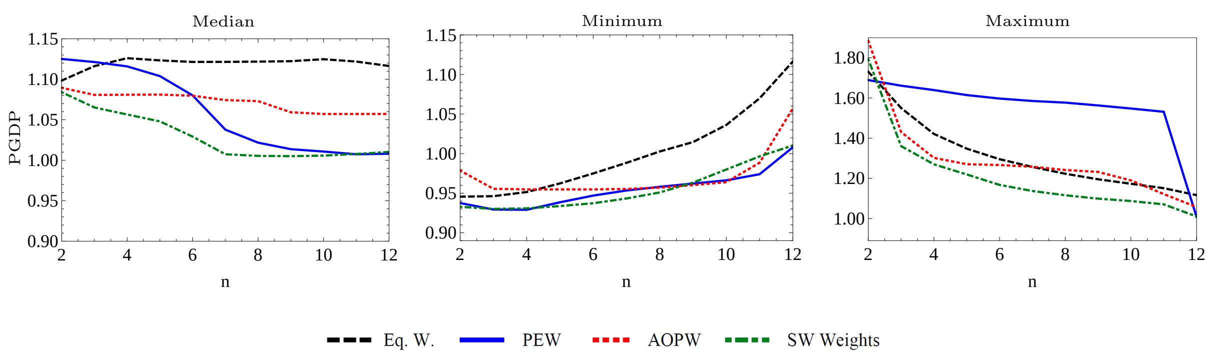

Figure A1: Median, minimum, and maximum relative RMSFE results for combinations of n different models. The parameter γ is set to zero. The results are relative to the RMSFE of the AO forecast.

Intercept Correction Results

	1970Q1-2014Q1		1983Q1-2007Q3		2007Q4-2014Q1	
Predictor	Rel. RMSFE	Bias	Rel. RMSFE	Bias	Rel. RMSFE	Bias
Benchmark						
AO	0.937	0.32†	1.262	0.49†	1.130	$-0.09^\dagger$
Direct Forecasts (relative RMSFE)						
DF CUR	1.562	1.81	2.703	3.06	1.493	2.17
DF GDP	1.513	1.62	2.553	3.20	1.258	1.89
DF Growth Gap	1.512	1.74	2.611	3.25	1.271	1.88
DF Output Gap	1.667	1.61	2.772	3.36	1.278	1.87
DF U. Gap	1.509	1.70	2.579	3.20	1.294	2.03
Phillips Curves (relative RMSFE)						
PC CUR	1.232	1.27	1.988	1.87	1.414	1.21†
PC GDP	1.197	1.13	1.895	2.11	1.217	1.03
PC Growth Gap	1.137	1.10	1.771	1.70	1.243	0.87†
PC Output Gap	1.309	1.17	2.036	2.27	1.231	1.13
PC U. Gap	1.179	1.11	1.843	1.82	1.352	1.14†
VAR ALL	1.296	1.09	2.073	1.99	1.229	0.91†
Univariate (relative RMSFE)						
AR(1)	1.297	1.24	2.141	2.48	1.227	1.28
AR(2)	1.139	1.05	1.806	1.93	1.196	1.00
AR(4)	1.208	1.22	1.917	2.20	1.217	1.15
ARMA(1, 1)	1.100	1.00	1.708	1.79	1.177	0.91†
ARMA(4, 4)	1.252	1.22	1.900	2.09	1.171	0.91†

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A3: Intercept correction results. The RMSFE results are presented relative to the uncorrected AO forecasts. Significance for the RMSFE results is only indicated for improvements over the benchmark. The † indicates a failure to reject the null hypothesis of unbiasedness at the 10% level.

Combined Forecast Results New Zealand

	1970Q1-2014Q1		1983Q1-2007Q3		2007Q4-2014Q1	
Predictor	Rel. RMSFE	Bias	Rel. RMSFE	Bias	Rel. RMSFE	Bias
Ouput Gap	1.073	1.068	$0.13^{\dagger}$	1.072	1.067	$0.09^{\dagger}$
U. Gap	0.942	0.938^{*}	$0.02^{\dagger}$	1.054	1.049	$-0.08^{\dagger}$
GDP Growth	0.989	0.984	$0.09^{\dagger}$	1.039	1.035	$0.05^{\dagger}$
NZ Output Gap	1.146	1.141	0.15	1.322	1.316	0.18
CUR	1.041	1.036	0.32	1.119	1.115	0.34
Comparisons						
AO	1.115	0.947	$0.12^{\dagger}$	–	–	–
Equal weights	1.004	1.000	$0.27^{\dagger}$	–	–	–

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A4: The RMSFEs are shown relative to the equal weights forecast. Significance for the RMSFE results is only indicated for improvements over the benchmark. The † indicates a failure to reject the null hypothesis of unbiasedness at the 10% level.

Appendix B

Proof We will 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

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

and the conditional problem

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

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 $\psi(\boldsymbol{w}) = \mathrm{E}[L(e_{c,T+h})|I_T]$. ∎